TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

A. A. ALBERT EINAR HILLE OSCAR ZARISKI

WITH THE COÖPERATION OF

RICHARD BRAUER NELSON DUNFORD W. K. FELLER
T. H. HILDEBRANDT S. C. KLEENE M. S. KNEBELMAN
R. E. LANGER SAUNDERS MacLANE OYSTEIN ORE
H. P. ROBERTSON J. J. STOKER D. J. STRUIK
GABOR SZEGÖ HASSLER WHITNEY R. L. WILDER

VOLUME 54

JULY TO DECEMBER, 1943

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1943

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

Chester C. Corbin Library Fund
Bound Jan. 1944

Composed, Printed and Bound by
The Collegiate Press
George Banta Publishing Company
Menasha, Wisconsin

QA
1
F001
v.54

TABLE OF CONTENTS

VOLUME 54, JULY TO DECEMBER, 1943

ALBERT, A. A. Quasigroups. I. 507

ALEXANDROFF, P. On homological situation properties of complexes and closed sets 286

BAER, R. A theory of crossed characters 103

BUSEMANN, H. On spaces in which two points determine a geodesic . . 171

CAMERON, R. H., and MARTIN, W. T. Infinite linear difference equations with arbitrary real spans and first degree coefficients 1

DECICCO, J., and KASNER, E. A generalized theory of dynamical trajectories 23

DUNFORD, N. Spectral theory. I. Convergence to projections 185

FELLER, W. Generalization of a probability limit theorem of Cramér . . 361

FELLER, W. The general form of the so-called law of the iterated logarithm 373

GREVILLE, T. N. E. Regularity of label-sequences under configuration transformations 403

HALL, M. Projective planes 229

HELSEL, R. G., and RADÓ, T. The transformation of double integrals . . 83

HOLTOM, C. Permanent configurations in the n-body problem 520

KASNER, E., and DECICCO, J. A generalized theory of dynamical trajectories 23

KOBER, H. On the approximation to integrable functions by integral functions 70

LORCH, E. R. The theory of analytic functions in normed abelian vector rings 414

MARTIN, W. T., and CAMERON, R. H. Infinite linear difference equations with arbitrary real spans and first degree coefficients 1

MUHLY, H. T. Independent integral bases and a characterization of regular surfaces 340

PURCELL, E. J. Flat space congruences of order one in $[n]$ 57

RADÓ, T., and HELSEL, R. G. The transformation of double integrals . . 83

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

SALEM, R. Sets of uniqueness and sets of multiplicity..... 218
SCHAEFFER, A. C. On the oscillation of differential transforms. III. Oscillations of the derivative of a function..... 278
SCHATTEN, R. On reflexive norms for the direct product of Banach spaces..... 498
SZÁSZ, O. On some trigonometric summability methods and Gibbs' phenomenon..... 483
WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large..... 426
ZYGMUND, A. A property of the zeros of Legendre polynomials..... 39

INFINITE LINEAR DIFFERENCE EQUATIONS WITH ARBITRARY REAL SPANS AND FIRST DEGREE COEFFICIENTS

BY

R. H. CAMERON AND W. T. MARTIN

1. Introduction. Finite difference equations with arbitrary real spans have been investigated by various mathematicians(1). In this paper we shall consider infinite difference equations with arbitrary real spans whose coefficients are either constants or linear functions of the independent variable.

We shall state at the outset three theorems which give some indication of the scope of our work. Theorems 1, 2 and 3 are special cases of Theorems 1a, 2a and 3a which will be proved in the body of the paper. We state these special cases explicitly because they are simpler and can be expressed in the usual terminology of difference equations. The three theorems are not independent since 1 and 2 together imply 3, and 2 is a special case of 3. Nevertheless, it seems worthwhile to quote 1 and 2 separately because of their greater simplicity and because they form the natural introduction to Theorem 3. They deal, respectively, with the case of constant coefficients, the case of all coefficients constant but one (which is linear), and the case of all linear coefficients.

THEOREM 1. *Let*

$$\sum_{n=1}^{\infty} a_n f(z - \lambda_n) = g(z) \tag{1.1}$$

be an infinite difference equation in which the λ_n are arbitrary real numbers and the a_n are complex numbers such that

$$\sum_{n=1}^{\infty} a_n e^{\lambda_n w} \tag{1.2}$$

converges absolutely to a non-vanishing function $A(w)$ in some strip $\alpha < \mathrm{Re}\ w < \beta$. Furthermore let $g(z)$ be analytic in some strip $a < \mathrm{Im}\ z < b$. If for some $\epsilon > 0$ inequalities

Presented to the Society, October 31, 1942; received by the editors June 24, 1942.

(1) A treatment of a very general nature has recently been given by T. Kitagawa, *On the theory of linear translatable functional equations and Cauchy's series*, Jap. J. Math. vol. 13 (1937) pp. 233–332. Kitagawa gives other references to the literature. The methods used in the treatment of the first theorem of the present paper are more closely related to those of S. Bochner, *Allgemeine lineare Differenzengleichungen mit asymptotisch konstanten Koeffizienten*, Math. Zeit. vol. 33 (1931) pp. 426–450, and W. T. Martin, *Linear difference equations with arbitrary real spans*, Acta Math. vol. 69 (1937) pp. 57–98. The other theorems use somewhat different methods.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

$$(1.3)\qquad \left.\begin{array}{ll} |g(x+iy)|\,e^{\beta' x} < K_{\alpha',\beta',a',b'} & \text{for } x<0 \\ |g(x+iy)|\,e^{\alpha' x} < K_{\alpha',\beta',a',b'} & \text{for } x>0 \end{array}\right\} a' < y < b'$$

hold for every α', β', a', b' *in*

$$(1.4)\qquad \alpha < \alpha' < \alpha + \epsilon, \quad \beta - \epsilon < \beta' < \beta, \quad a < a' < b' < b,$$

then it follows that (1.1) *has one and only one solution* $f(z)$ *of the same character, that is, analytic in* $a < \operatorname{Im} z < b$ *and satisfying*

$$(1.5)\qquad \left.\begin{array}{ll} |f(x+iy)|\,e^{\beta' x} < K'_{\alpha',\beta',a',b'} & \text{for } x<0 \\ |f(x+iy)|\,e^{\alpha' x} < K'_{\alpha',\beta',a',b'} & \text{for } x>0 \end{array}\right\} a' < y < b'$$

for every α', β', a', b' *in* (1.4). *Moreover* $f(z)$ *is given explicitly by the formula*

$$f(z) = \sum_{n=1}^{\infty} a_n' g(z - \lambda_n'), \qquad a < \operatorname{Im} z < b,$$

where

$$\frac{1}{A(w)} = \sum_{n=1}^{\infty} a_n' e^{\lambda_n' w}, \qquad \alpha < \operatorname{Re} w < \beta\ (^2).$$

THEOREM 2. *Let*

$$(1.6)\qquad zf(z) = \sum_{n=1}^{\infty} b_n f(z - \lambda_n) + g(z)$$

be a difference equation in which the λ_n *are arbitrary real numbers and the* b_n *are complex numbers such that*

$$(1.7)\qquad \sum_{n=1}^{\infty} b_n e^{\lambda_n w}$$

converges absolutely to a function $B(w)$ *in some strip* $\alpha < \operatorname{Re} w < \beta$ *of the auxiliary* w*-plane. Let* y_0 *be the imaginary part of the constant term of* (1.7) *(zero if it has none). Then if* $g(z)$ *is analytic and satisfies* (1.3) *where the strip* $a < \operatorname{Im} z < b$ *does not contain the line* $\operatorname{Im} z = y_0$, *it follows that* (1.6) *has a unique solution* $f(z)$ *of the same character, that is, analytic in* $a < \operatorname{Im} z < b$ *and satisfying* (1.5).

THEOREM 3. *Let*

$$(1.8)\qquad \sum_{n=1}^{\infty} (a_n z + b_n) f(z - \lambda_n) = g(z)$$

(2) Throughout this paper two complex planes will enter into consideration, a z-plane ($z = x + iy$) in which the independent variable in our functional equation ranges, and an auxiliary w-plane ($w = u + iv$) in which we consider an associated function which governs the growth of the solutions in the z-plane.

be a difference equation involving arbitrary real spans and complex linear coefficients such that

$$(1.9) \qquad \sum_{n=1}^{\infty} a_n e^{\lambda_n w} \quad \text{and} \quad \sum_{n=1}^{\infty} (b_n + a_n\lambda_n) e^{\lambda_n w}$$

converge absolutely to functions $A(w)$ and $\tilde{B}(w)$ in the strip $\alpha < \mathrm{Re}\, w < \beta$, the function $A(w)$ being non-vanishing in the strip. Then

$$y_0 = -\,\mathrm{Im}\left\{\lim_{T\to\infty} \frac{1}{Ti} \int_{\gamma}^{\gamma+Ti} \frac{\tilde{B}(w)}{A(w)}\, dw\right\}$$

will exist and be independent of γ $(\alpha < \gamma < \beta)$; and if $g(z)$ is analytic and satisfies (1.3) *where the strip $a < \mathrm{Im}\, z < b$ does not contain the line $\mathrm{Im}\, z = y_0$, it follows that* (1.8) *has a unique solution $f(z)$ of the same character, that is, analytic in and satisfying* (1.5).

The difference equations which we have just stated can all be expressed in terms of Stieltjes integrals. The difference equations are therefore special cases of certain Stieltjes integral equations, which are dealt with in Theorems 1a, 2a and 3a. These theorems are proved by means of the theory of Fourier-Stieltjes transforms. Theorem 1a which depends on Pitt's lemma(3) is in a certain sense not a true Stieltjes integral theorem since the Stieltjes integral involved is restricted to have no singular part, and the equation is thus really only an infinite integro-difference equation. The same is true of Theorem 3a insofar as it depends upon Theorem 1a. No such restriction is required in Theorem 2a, and it is therefore strictly a Stieltjes integral equation theorem.

If we form a right-continuous step function $a(\lambda)$ with $a(0) = 0$ and with jumps of a_n at points λ_n,

$$(1.10) \qquad a(\lambda) = \begin{cases} \sum_{0<\lambda_n<\lambda} a_n, & \lambda \geqq 0, \\ -\sum_{\lambda<\lambda_n<0} a_n, & \lambda < 0, \end{cases}$$

then the condition that the series (1.2) be absolutely convergent in $\alpha < \mathrm{Re}\, w < \beta$ can be expressed in an alternative but equivalent form.

LEMMA 1.1. *A necessary and sufficient condition that the series* (1.1) *be absolutely convergent in $\alpha < \mathrm{Re}\, w < \beta$ is that*

(3) H. R. Pitt, *Mercerian theorems*, Proc. Cambridge Philos. Soc. vol. 34 (1938) pp. 510–520. See Lemma, p. 513. See also I. Gelfand, *Über absolut konvergente trigonometrische Reihen und Integrale*, Rec. Math. (Mat. Sbornik) N.S. vol. 9 (1941) pp. 51–66. Equation (3.4) of Theorem 1a has been solved by Pitt on the real line in the above paper and in *General Mercerian theorems* (II), Proc. London Math. Soc. vol. 47 (1942) pp. 248–267.

$$\int_{-\infty}^{0} e^{\alpha'\lambda} \left| da(\lambda) \right| + \int_{0}^{\infty} e^{\beta'\lambda} \left| da(\lambda) \right| < \infty \tag{1.11}$$

for every α', β' *in* $\alpha < \alpha' < \beta' < \beta$.

Also the sum-function $A(w)$ of (1.2) is given by $A(w) = \int_{-\infty}^{\infty} e^{\lambda w} da(\lambda)$, $\alpha < \operatorname{Re} w < \beta$.

The growth condition (1.3) can also be expressed in an alternative and equivalent form, as we shall show.

LEMMA 1.2. *Let* $g(z)$ *be analytic in a strip* $a < y < b$, $-\infty < x < \infty$, *and let* α, β, ϵ *be given real numbers with* $\epsilon > 0$. *Then a necessary and sufficient condition that* (1.3) *shall hold for every* α', β', a', b' *in* (1.4) *is that*

$$\sup_{a' < y < b'} \left[\int_{-\infty}^{0} e^{\beta' x} \left| g(x + iy) \right| dx + \int_{0}^{\infty} e^{\alpha' x} \left| g(x + iy) \right| dx \right] < \infty \tag{1.12}$$

hold for every α', β', a', b' *in* (1.4).

Lemma 1.2 is an immediate consequence of the following property of analytic functions, which is based upon the Phragmén-Lindelöf principle:

Let $h(z)$ *be analytic in the half-strip*

$$c \leqq \operatorname{Im} z \leqq d, \qquad 0 \leqq \operatorname{Re} z < \infty, \tag{1.13}$$

and let

$$\int_{0}^{\infty} \left| h(x + iy) \right| dx \leqq M, \qquad c \leqq y \leqq d. \tag{1.14}$$

Then for every strip (c', d') *interior to* (c, d) *we have*

$$\left| h(x + iy) \right| < M_{c',d'}, \qquad c' < y < d', \ 0 \leqq x. \tag{1.15}$$

Proof. Let z be any point in (1.13) and define

$$H(z) = \int_{ic}^{z} h(t)\,dt. \tag{1.16}$$

The function $H(z)$ so defined is analytic in (1.13) and is independent of the path as long as the path lies in (1.13). On taking the path to be the sum of two straight-line segments $(ic, i \operatorname{Im} z)$ and $(i \operatorname{Im} z, z)$ we find

$$\left| H(z) \right| \leqq \int_{c}^{d} \left| h(iy) \right| dy + \int_{0}^{x} \left| h(u + iy) \right| du \leqq M_1 + M.$$

Thus $H(z)$ is bounded in (1.13). Moreover by (1.14) we see that $H(z)$ approaches a limit as z approaches infinity along any one horizontal half-line in $c \leqq \operatorname{Im} z \leqq d$. Thus by a well known form of the Phragmén-Lindelöf theorem $H(z)$ approaches one limit value uniformly in the strip (1.13).

Next if $c<c'<d'<d$ and δ is a sufficiently small positive number, then for $c'<y<d'$ and $\delta<x$,

$$h(x+iy)=\frac{1}{2\pi i}\int_{|t-z|=\delta}\frac{H(t)}{(t-z)^2}dt$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{H(z+\delta e^{i\phi})}{\delta e^{i\phi}}d\phi\to(\text{const.})\frac{1}{\delta}\int_0^{2\pi}e^{-i\phi}d\phi=0,$$

as $x\to\infty$, the convergence being uniform in $c'<y<d'$. Thus (1.15) holds.

2. **Norms and their properties.** Before giving the necessary lemmas and definitions for our work on integral equations we will make a few remarks which indicate the motivation for the remainder of the section. In order to include the previously stated theorems on difference equations under our theorems on integral equations, as well as to have more generality in the integral equations themselves, we must dea' with functions $p(\lambda)$ of bounded variation on every finite interval but not necessarily on the whole infinite interval. Thus the expression $\int_{-\infty}^{\infty}|dp(\lambda)|$ need not necessarily be finite for the functions $p(\lambda)$ which we consider. We shall, however, impose a certain condition on the behaviour of $p(\lambda)$ which will enable us to form Radon-Stieltjes integrals(4) of the form $\int_{-\infty}^{\infty}h(\lambda)dp(\lambda)$ for certain classes of functions $h(\lambda)$. In fact we shall define a norm $N_1(p)$ for right-continuous functions $p(\lambda)$ of bounded variation on every finite interval and we shall work with functions $p(\lambda)$ of finite norm, $N_1(p)<\infty$. If $p(\lambda)$ and $q(\lambda)$ are two such functions, each of finite norm, then the convolution $\int_{-\infty}^{\infty}q(\lambda-\mu)dp(\mu)$ need not necessarily exist, but nevertheless a certain modified convolution can be shown to exist, namely $p\times q=\int_{-\infty}^{\infty}[q(\lambda-\mu)-q(-\mu)dp(\mu)=\int_{-\infty}^{\infty}[p(\lambda-\mu)-p(-\mu)]dq(\mu)$. And this modified convolution serves the usual purposes required for the ordinary convolution. For example, if $P(w)$ and $Q(w)$ are the Fourier-Stieltjes transform of $p(\lambda)$ and $q(\lambda)$, respectively, $P(w)=\int_{-\infty}^{\infty}e^{\lambda w}dp(\lambda)$, $Q(w)=\int_{-\infty}^{\infty}e^{\lambda w}dq(\lambda)$ then the product $R(w)=P(w)Q(w)$ is the Fourier-Stieltjes transform of the modified convolution $r(\lambda)\equiv p(\lambda)\times q(\lambda)$.

We will also define a (weak) norm $N(f)$ for analytic functions $f(z)$ in a strip $a<\text{Im } z<b$, this norm being related to our N_1-norm, and we shall show in particular that the ordinary convolution $\int_{-\infty}^{\infty}f(z-\lambda)dp(\lambda)$ exists and yields a function analytic in $a<\text{Im } z<b$ and of finite N-norm at most $N(f)N_1(p)$, whenever $N(f)$ and $N_1(p)$ are finite. The fact that the N-norm is a weak norm will be useful in our functional equations since it will enable us to treat

(4) The integrals which occur in this paper are understood to be Lebesgue-Stieltjes (Radon) integrals. See, for example, Saks, *Theory of the integral*, 2d revised edition, Warsaw-Lemberg, 1937, pp. 19 and 67. We note that the familiar principles of monotone and dominated convergence in the Lebesgue theory as well as the symmetric Fubini theorem are also valid in this theory; see pp. 28, 29, and 77. In this paper we shall not admit $\pm\infty$ as members of our number system; that is, existence implies finiteness.

analytic functions of finite exponential growth in a strip, the rate of growth being permitted to increase as the rate of growth of the function $p(\lambda)$ of bounded variation decreases.

We now proceed with the necessary definitions and lemmas. Let two positive continuous functions $E(x)$ and $E_1(x)$ satisfying

$$E(x+y) \leqq E(x)E_1(y), \quad E_1(x+y) \leqq E_1(x)E_1(y), \tag{2.1}$$

for all x and y, be given. These functions will be designated throughout the paper by this same notation. In practice, $E(x)$ and $E_1(x)$ will usually be

$$E(x) = \begin{cases} e^{\alpha x}, & x \geqq 0, \\ e^{\beta x}, & x < 0, \end{cases} \qquad E_1(x) = \begin{cases} e^{\beta x}, & x \geqq 0, \\ e^{\alpha x}, & x < 0, \end{cases}$$

with $\alpha < \beta$. The reader can readily verify that these functions satisfy (2.1). In terms of definite preassigned functions $E(x)$ and $E_1(x)$, we define the following two norms:

DEFINITION 1. *Let $p(\lambda)$ be a right-continuous function of bounded variation on every finite interval. Then we define the norm*

$$N_1(p) = \int_{-\infty}^{\infty} E_1(\lambda) \, | \, dp(\lambda) \, | . \tag{2.2}$$

DEFINITION 2. *Let $f(z) = f(x+iy)$ be defined in a strip $a < y < b$. Then we define the norms*

$$N(f; y) = \int_{-\infty}^{\infty} E(x) \, | \, f(x+iy) \, | \, dx \tag{2.3a}$$

and

$$N(f) = \sup_{a<y<b} N(f; y). \tag{2.3b}$$

In practice the N-norm will be used for functions which are analytic in a strip. Nevertheless we use this general definition which applies to any function of a complex variable, since on occasions we will use the finiteness of the norm to aid in proving the analyticity.

We shall use the term "of finite norm" with reference to both functions of a complex variable and functions of bounded variation, with the understanding that the appropriate norm will be used in each case. While there is a possible overlapping of these types of functions, in practice no confusion will arise from the use of the term "of finite norm."

We shall use the following properties of the norm, where the existence (finiteness) of the right side implies the existence of the left.

LEMMA 2.1.

$$N_1(p_1 + p_2) \leqq N_1(p_1) + N_1(p_2), \tag{2.4}$$

$$N(f_1 + f_2; y) \leqq N(f_1; y) + N(f_2; y), \tag{2.5}$$

$$N(f_1 + f_2) \leqq N(f_1) + N(f_2), \tag{2.5a}$$

$$N\left[\int_{-\infty}^{\infty} f(\cdot - \lambda)dp(\lambda)\right] \leqq N_1(p)N(f), \tag{2.6}$$

$$N\left(\frac{f}{\cdot - c}; y\right) \leqq \frac{1}{|\operatorname{Im}(z - c)|} N(f; y). \tag{2.7}$$

The inequalities (2.4), (2.5), (2.5a), (2.7) follow immediately from the definitions; so we proceed to the proof of (2.6). Obviously it is sufficient to prove

$$N\left[\int_{-\infty}^{\infty} f(\cdot - \lambda)dp(\lambda); y\right] \leqq N_1(p)N(f; y) \tag{2.8}$$

and we may without loss of generality assume that $y=0$. Then we have

$$\begin{aligned} N\left[\int_{-\infty}^{\infty} f(\cdot - \lambda)dp(\lambda); 0\right] &= \int_{-\infty}^{\infty} E(x)\left|\int_{-\infty}^{\infty} f(x - \lambda)dp(\lambda)\right|dx \\ &\leqq \int_{-\infty}^{\infty} E(x)dx \int_{-\infty}^{\infty} |f(x - \lambda)|\,|dp(\lambda)| \\ &= \int_{-\infty}^{\infty} |dp(\lambda)| \int_{-\infty}^{\infty} E(x)\,|f(x - \lambda)|\,dx \\ &\leqq \int_{-\infty}^{\infty} |dp(\lambda)| \int_{-\infty}^{\infty} E_1(\lambda)E(x - \lambda)\,|f(x - \lambda)|\,dx \\ &= \int_{-\infty}^{\infty} E_1(\lambda)\,|dp(\lambda)| \int_{-\infty}^{\infty} E(x)\,|f(x)|\,dx \\ &= N_1(p)N(f; 0). \end{aligned}$$

(It is clear from this argument that $\int_{-\infty}^{\infty}|f(x-\lambda)|\cdot|dp(\lambda)|$ and hence $\int_{-\infty}^{\infty}f(x-\lambda)dp(\lambda)$ must exist for almost all x when $N_1(p)$ and $N(f; 0)$ exist.)

LEMMA 2.2. *Let $f(z)$ be analytic in $a<\operatorname{Im} z<b$, and let $\delta>0$. Then for any point $z=x+iy$ in the strip $a+2\delta<\operatorname{Im} z<b-2\delta$ we have*

$$|f(z)| \leqq \frac{2}{\pi\delta} N(f) \cdot \sup_{|x-\xi|<\delta} [E(\xi)]^{-1}. \tag{2.9}$$

This follows from the fact that $f(z)$ is a harmonic function of x and y. Thus

$$f(z) = \frac{1}{\pi\delta^2}\iint_{|z-w|<\delta} f(u+iv)dudv, \qquad w = u+iv,$$

and

$$\begin{aligned}
\pi\delta^2 |f(z)| &\leqq \int_{y-\delta}^{y+\delta}\int_{x-\delta}^{x+\delta} |f(u+iv)|\, dudv \\
&\leqq \int_{y-\delta}^{y+\delta}\int_{x-\delta}^{x+\delta} |f(u+iv)|\,[E(u)]^{-1}E(u)dudv \\
&\leqq \sup_{|x-\xi|<\delta}[E(\xi)]^{-1}\int_{y-\delta}^{y+\delta}\int_{-\infty}^{\infty} |f(u+iv)|\,E(u)dudv \\
&\leqq \sup_{|x-\xi|<\delta}[E(\xi)]^{-1}\int_{y-\delta}^{y+\delta} N(f)dv \\
&= 2\delta \sup_{|x-\xi|<\delta}[E(\xi)]^{-1}\cdot N(f).
\end{aligned}$$

LEMMA 2.3. *If $f(z)$ is analytic in the strip $a<\operatorname{Im} z<b$ and if $N(f)$ and $N_1(p)$ are both finite, then $\int_{-\infty}^{\infty} f(z-\lambda)dp(\lambda)$ is analytic in the strip $a<\operatorname{Im} z<b$.*

To prove this, we shall show that

$$\int_0^{\infty} f(z-\lambda)dp(\lambda) = \lim_{M\to\infty}\int_0^{M} f(z-\lambda)dp(\lambda) \tag{2.10}$$

uniformly in z in the rectangle R_δ: $a+2\delta<\operatorname{Im} z<b-2\delta$, $A+\delta<\operatorname{Re} z<B-\delta$, where A and B are arbitrary. Now by (2.9), (2.6), (2.2) we have in R_δ:

$$\begin{aligned}
\left|\int_{M_1}^{M_2} f(z-\lambda)dp(\lambda)\right| &\leqq \frac{2}{\pi\delta} N\left[\int_{M_1}^{M_2} f(\cdot-\lambda)dp(\lambda)\right]\cdot \sup_{|z-\xi|<\delta}[E(\xi)]^{-1} \\
&\leqq \frac{2}{\pi\delta} N(f)\cdot\left[\int_{M_1}^{M_2} E_1(\lambda)\,|\,dp(\lambda)|\right]\cdot \sup_{A<\xi<B}[E(\xi)]^{-1},
\end{aligned}$$

since we can consider the finite integral as an infinite integral with $p(\lambda)$ replaced by a function which is constant outside the finite interval. Thus $\lim_{M_1,M_2\to\infty}\int_{M_1}^{M_2} f(z-\lambda)dp(\lambda)=0$ uniformly in R_δ, and the uniform convergence in (2.10) is established. Since the finite integral is obviously analytic, the same is true of the infinite and also of the two-way infinite integral.

LEMMA 2.4. *Let $\{f_n(z)\}$ be a sequence of analytic functions in the strip $a<\operatorname{Im} z<b$ such that*

$$\sum_{n=1}^{\infty} N(f_n) < \infty. \tag{2.11}$$

Then the series

$$f(z) = \sum_{n=1}^{\infty} f_n(z) \tag{2.12}$$

converges absolutely and uniformly in every closed finite region contained in the strip $a < \operatorname{Im} z < b$ *and represents an analytic function* $f(z)$ *of finite norm in the strip. Moreover*

$$N(f) \leqq \sum_{n=1}^{\infty} N(f_n), \tag{2.13}$$

and if $N_1(f) < \infty$, *then*

$$\int_{-\infty}^{\infty} \left[\sum_{n=1}^{\infty} f_n(z - \lambda) \right] d\rho(\lambda) = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} f_n(z - \lambda) d\rho(\lambda) \tag{2.14}$$

where the right member converges uniformly in every closed finite subregion of the strip.

The absolute and uniform convergence of (2.12) in closed finite subregions of the strip follows immediately from (2.9) and (2.11). Thus $f(z)$ is analytic in the strip and we have

$$\begin{aligned} N(f; y) &= \int_{-\infty}^{\infty} E(x) \, | f(x + iy) | \, dx = \int_{-\infty}^{\infty} E(x) \left| \sum_{n=1}^{\infty} f_n(x + iy) \right| dx \\ &\leqq \int_{-\infty}^{\infty} E(x) \sum_{n=1}^{\infty} | f_n(x + iy) | \, dx = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} E(x) \, | f_n(x + iy) | \, dx \\ &= \sum_{n=1}^{\infty} N(f_n; y). \end{aligned}$$

Moreover if $h(z) = \int_{-\infty}^{\infty} f(z-\lambda) d\rho(\lambda)$ and $h_n(z) = \int_{-\infty}^{\infty} f_n(z-\lambda) d\rho(\lambda)$, and $a + 2\delta < \operatorname{Im} z < b - 2\delta$, then

$$\left| h(z) - \sum_{n=1}^{M-1} h_n(z) \right| = \left| \int_{-\infty}^{\infty} \sum_{n=M}^{\infty} f_n(z - \lambda) d\rho(\lambda) \right| \tag{2.15}$$

[by (2.9)] $$\leqq \frac{2}{\pi\delta} N\left(\int_{-\infty}^{\infty} \sum_{n=M}^{\infty} f_n(\cdot - \lambda) d\rho(\lambda) \right) \sup_{|z-\xi|<\delta} [E(\xi)]^{-1}$$

[by (2.6)] $$\leqq \frac{2}{\pi\delta} N_1(\rho) N\left(\sum_{n=M}^{\infty} f_n \right) \sup_{|z-\xi|<\delta} [E(\xi)]^{-1}$$

[by (2.13)] $$\leqq \frac{2}{\pi\delta} N_1(\rho) \left[\sum_{n=M}^{\infty} N(f_n) \right] \sup_{|z-\xi|<\delta} [E(\xi)]^{-1}$$

[by (2.11)] $$\to 0 \text{ as } M \to \infty.$$

This yields (2.14).

LEMMA 2.5. *Let $p(\lambda)$ be a right-continuous function of bounded variation on every finite interval and of finite norm. For any values of λ and μ we have*

$$
\begin{aligned}
| p(\lambda - \mu) - p(-\mu) | &\leqq \left\{ \sup_{\nu \in [-\mu, \lambda - \mu]} \frac{1}{E_1(\nu)} \right\} N_1(p) \\
&\leqq \left\{ \sup_{\nu \in [0, \lambda]} \frac{1}{E_1(\nu)} \right\} E_1(\mu) N_1(p)
\end{aligned} \tag{2.16}
$$

where $[\alpha_1, \alpha_2]$ denotes the closed interval with endpoints α_1 and α_2 ($\alpha_1 \geqq \alpha_2$, $\alpha_1 < \alpha_2$).

Proof.

$$
\begin{aligned}
| p(\lambda - \mu) - p(-\mu) | &= \left| \int_{-\mu}^{\lambda - \mu} dp(s) \right| \leqq \left| \int_{-\mu}^{\lambda - \mu} \frac{1}{E_1(s)} E_1(s) dp(s) \right| \\
&\leqq \left\{ \sup_{\nu \in [-\mu, \lambda - \mu]} \frac{1}{E_1(\nu)} \right\} N_1(p).
\end{aligned}
$$

The second inequality in (2.16) follows from (2.1). Thus

$$1/E_1(\nu - \mu) \leqq E_1(\mu)/E_1(\nu).$$

LEMMA 2.6. *If $p(\lambda)$ and $q(\lambda)$ are of bounded variation on every finite interval and of finite norm then the modified convolution*

$$p \times q \equiv \int_{-\infty}^{\infty} [p(\lambda - \mu) - p(-\mu)] dq(\mu) \tag{2.17}$$

exists and is of bounded variation on every finite interval and is of finite norm,

$$N_1\left(\int_{-\infty}^{\infty} [p(\cdot - \mu) - p(-\mu)] dq(\mu) \right) \leqq N_1(p) N_1(q). \tag{2.18}$$

Proof. By (2.16)

$$
\begin{aligned}
&\int_{-\infty}^{\infty} | p(\lambda - \mu) - p(-\mu) | \, | dq(\mu) | \\
&\qquad \leqq \left\{ \sup_{\nu \in [0, \lambda]} \frac{1}{E_1(\nu)} \right\} N_1(p) \int_{-\infty}^{\infty} E_1(\mu) \, | dq(\mu) | \\
&\qquad = \left\{ \sup_{\nu \in [0, \lambda]} \frac{1}{E_1(\nu)} \right\} N_1(p) N_1(q).
\end{aligned}
$$

Thus the integral in (2.17) exists (is finite). If p and q are both increasing functions then the modified convolution is increasing. It readily follows that (2.17) is of bounded variation in the general case. For (2.18) we have

$$
\begin{aligned}
N_1\Big\{\int_{-\infty}^{\infty}[p(\lambda-\mu)-p(-\mu)]dq(\mu)\Big\}
&= \int_{-\infty}^{\infty}E_1(\lambda)\Big|d_\lambda\int_{-\infty}^{\infty}(p(\lambda-\mu)-p(-\mu)]dq(\mu)\Big| \\
&\leqq \int_{-\infty}^{\infty}E_1(\lambda)d_\lambda\Big\{\int_{-\infty}^{\infty}\mathrm{Var}\,[p(\lambda-\mu)-p(-\mu)]\,|\,dq(\mu)\,|\Big\} \\
&= \int_{-\infty}^{\infty}|\,dq(\mu)\,|\int_{-\infty}^{\infty}E_1(\lambda)\,|\,d_\lambda[p(\lambda-\mu)-p(-\mu)]\,| \\
&= \int_{-\infty}^{\infty}|\,dq(\mu)\,|\int_{-\infty}^{\infty}E_1(\mu+\nu)\,|\,dp(\nu)\,| \\
&\leqq \int_{-\infty}^{\infty}|\,dq(\mu)\,|\int_{-\infty}^{\infty}E_1(\mu)E_1(\nu)\,|\,dp(\nu)\,| \\
&= N_1(p)N_1(q);
\end{aligned}
$$

which proves (2.18)[5].

In some places it will be useful to have our functions normalized in the following way. We call $p(\lambda)$ *normalized* if (i) $p(\lambda)$ is right-continuous and (ii) $p(0)=0$.

COROLLARY TO LEMMA 2.6. *If p is normalized then the convolution $p\times q$ is normalized. If p and q are normalized then the convolution is symmetric: $p\times q=q\times p$.*

LEMMA 2.7. *Let $\{p_n(\lambda)\}$ be a sequence of normalized functions, each of bounded variation on every finite interval, such that*

$$\sum_{n=1}^{\infty}N_1(p_n)<\infty. \tag{2.19}$$

Then on every finite interval the series

$$p(\lambda)=\sum_{1}^{\infty}p_n(\lambda) \tag{2.20}$$

converges uniformly to a function of bounded variation, and is a function of finite norm. Moreover

$$N_1(p)\leqq\sum_{n=1}^{\infty}N_1(p_n). \tag{2.21}$$

---

[5] To see that the interchange of order of integration just made is justifiable compare *An unsymmetric Fubini theorem* by R. H. Cameron and W. T. Martin, Bull. Amer. Math. Soc. vol. 47 (1941) pp. 121–125. Throughout the paper we shall frequently make use of the result to permit an interchange of order of integration of two integrals of this form.

Furthermore, if $N_1(q) < \infty$, *then*

$$\int_{-\infty}^{\infty}\left[\sum_{n=1}^{\infty}\{p_n(\lambda-\mu)-p_n(-\mu)\}\right]dq(\mu) = \sum_{n=1}^{\infty}\int_{-\infty}^{\infty}\{p_n(\lambda-\mu)-p_n(-\mu)\}dq(\mu), \tag{2.22}$$

where in every finite interval the right member converges uniformly to a function of bounded variation.

The uniform convergence of (2.20) follows from Lemma 2.5. That the sum is of bounded variation can be seen by separating into increasing parts. Thus if

$$p_n^{\pm}(\lambda) = \frac{1}{2}\int_0^{\lambda} |\,dp_n(\mu)\,| \pm p_n(\lambda)$$

then

$$N_1(p_n^{\pm}) \leqq N_1(p_n)$$

and

$$p(\lambda) = p^{+}(\lambda) - p^{-}(\lambda) \equiv \sum_{n=1}^{\infty} p_n^{+}(\lambda) - \sum_{n=1}^{\infty} p_n^{-}(\lambda),$$

so that $p(\lambda)$ is of bounded variation on every finite interval.

Next, (2.21) follows for increasing parts:

$$\begin{aligned}\int_{-M}^{M} E_1(\lambda)d\left[p^{\pm}(\lambda) - \sum_{n=1}^{K} p_n^{\pm}(\lambda)\right] &\leqq \left[\max_{-M\leq\lambda\leq M} E_1(\lambda)\right]\int_{-M}^{M} d\left[p^{\pm}(\lambda) - \sum_{1}^{K} p_n^{\pm}(\lambda)\right] \\ &= \left[\max_{-M\leq\lambda\leq M} E_1(\lambda)\right]\left\{\left[p^{\pm}(M) - \sum_{1}^{K} p_n^{\pm}(M)\right]\right. \\ &\qquad \left. - \left[p^{\pm}(-M) - \sum_{1}^{K} p_n^{\pm}(-M)\right]\right\} \\ &\to 0 \text{ as } K \to \infty, \text{ for fixed } M.\end{aligned}$$

This yields

$$\int_{-M}^{M} E_1(\lambda)dp^{\pm}(\lambda) = \sum_{n=1}^{\infty}\int_{-M}^{M} E_1(\lambda)dp_n^{\pm}(\lambda) \leqq \sum_{n=1}^{\infty} N_1(p_n^{\pm}).$$

Letting M approach infinity we obtain (2.21) for increasing parts. For the general case,

$$N_1(p) = N_1(p^+ - p^-) \leqq N_1(p^+) + N_1(p^-)$$
$$\leqq \sum_1^\infty N_1(p_n^+) + \sum_1^\infty N_1(p_n^-) \leqq 2\sum_1^\infty N_1(p_n),$$

and

$$N_1\Big(\sum_{n=1}^\infty p_n\Big) = N_1\Big(\sum_{n=1}^K p_n + \sum_{n=k+1}^\infty p_n\Big) \leqq N_1\Big(\sum_1^K p_n\Big) + N_1\Big(\sum_{k+1}^\infty p_n\Big)$$
$$\leqq \sum_1^K N_1(p_n) + 2\sum_{k+1}^\infty N_1(p_n) \to \sum_1^\infty N_1(p_n) \text{ as } K \to \infty.$$

This yields (2.21) in general.

For (2.22) we use Lemma 2.5 and condition (2.19) and obtain

$$\int_{-\infty}^\infty \sum_{n=1}^\infty |\, p_n(\lambda - \mu) - p_n(-\mu)\,|\,|\,dq(\mu)\,|$$
$$\leqq \Big\{ \sup_{\nu \in [0,\lambda]} \frac{1}{E_1(\nu)} \Big\} \int_{-\infty}^\infty E_1(\mu) \Big[\sum_{n=1}^\infty N_1(p_n) \Big] |\, dq(\mu)\,|$$
$$= \Big\{ \sup_{\nu \in [0,\lambda]} \frac{1}{E_1(\nu)} \Big\} \Big[\sum_1^\infty N_1(p_n) \Big] N_1(q).$$

This enables us to apply the principle of dominated convergence to obtain (2.22). That the right member of (2.22) converges uniformly in every finite interval to a function of bounded variation follows immediately from Lemma 2.6 and the earlier part of the present lemma already proved.

3. The difference equation with constant coefficients. We shall now treat equation (1.1). We obtain a slightly more general result than Theorem 1 in terms of an integral equation.

THEOREM 1a. *Let $p(\lambda)$ be a right-continuous complex-valued function of bounded variation on every finite interval, whose singular part is zero; let p be of finite norm,*

$$(3.1) \qquad N_1(p) \equiv \int_{-\infty}^0 e^{\alpha\lambda} |\, dp(\lambda)\,| + \int_0^\infty e^{\beta\lambda} |\, dp(\lambda)\,| < \infty,$$

and let p be such that the (analytic) function

$$(3.2) \qquad P(w) = \int_{-\infty}^\infty e^{\lambda w} dp(\lambda)$$

is non-vanishing in the strip $\alpha < \mathrm{Re}\ w < \beta$. Let $g(z)$ be a function analytic in the strip $a < \mathrm{Im}\ z < b$ and of finite N-norm,

$$(3.3) \quad N(g) \equiv \sup_{a<y<b} \Big[\int_{-\infty}^0 e^{\beta x} |\, g(x + iy)\,| \, dx + \int_0^\infty e^{\alpha x} |\, g(x + iy)\,| \, dx \Big] < \infty.$$

Then the equation

$$\int_{-\infty}^{\infty} f(z-\lambda)dp(\lambda) = g(z) \tag{3.4}$$

has one and only one solution $f(z)$ which is analytic in the strip $a<\text{Im } z<b$ and which is of finite norm $N(f)<\infty$.

For the proof let us notice that since $P(w)$ is analytic and non-vanishing in the strip $\alpha<\text{Re } w<\beta$ and since (3.1) holds it follows from a result of Pitt (loc. cit.) that $1/P(w)$ is expressible in the form

$$\frac{1}{P(w)} = \int_{-\infty}^{\infty} e^{\lambda w} dp^*(\lambda), \qquad \alpha < \text{Re } w < \beta, \tag{3.5}$$

where $p^*(\lambda)$ is a function of bounded variation on every finite interval with finite norm. Also(6) since

$$\begin{aligned} 1 = \frac{P(w)}{P(w)} &= \int_{-\infty}^{\infty} e^{\lambda w} dp(\lambda) \int_{-\infty}^{\infty} e^{\mu w} dp^*(\mu) \\ &= \int_{-\infty}^{\infty} dp^*(\mu) \int_{-\infty}^{\infty} e^{\nu w} d_\nu p(\nu-\mu) \\ &= \int_{-\infty}^{\infty} dp^*(\mu) \int_{-\infty}^{\infty} e^{\nu w} d_\nu [p(\nu-\mu) - p(-\mu)] \\ &= \int_{-\infty}^{\infty} e^{\nu w} d_\nu \int_{-\infty}^{\infty} [p(\nu-\mu) - p(-\mu)] dp^*(\mu), \qquad \alpha < \text{Re } w < \beta, \end{aligned} \tag{3.6}$$

it follows that $\int_{-\infty}^{\infty}[p(\nu-\mu)-p(-\mu)]dp^*(\mu)$ is a function with a unit jump at the origin and constant elsewhere. Next we form the function

$$f(z) = \int_{-\infty}^{\infty} g(z-\mu)dp^*(\mu) \tag{3.7}$$

which by Lemmas 2.3 and 2.1 is analytic in the strip $a<\text{Im } z<b$, and of finite norm. Moreover

$$\begin{aligned} \int_{-\infty}^{\infty} f(z-\lambda)dp(\lambda) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(z-\lambda-\mu)dp^*(\mu)dp(\lambda) \\ &= \int_{-\infty}^{\infty} g(z-\nu)d_\nu \int_{-\infty}^{\infty} [p(\nu-\mu) - p(-\mu)]dp^*(\mu) \\ &= g(z). \end{aligned} \tag{3.8}$$

Thus $f(z)$ satisfies the equation (3.4).

(6) See Footnote 4.

For uniqueness let us assume that $\bar{f}$ is any solution of (3.4) of finite norm. Then

$$f(z) = \int_{-\infty}^{\infty} g(z-\mu)dp^*(\mu) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \bar{f}(z-\mu-\lambda)dp(\lambda)\right]dp^*(\mu) = \bar{f}(z);$$

that is, $f \equiv \bar{f}$.

This concludes the proof of Theorem 1a, and also shows that the solution $f(z)$ is given explicitly by (3.7).

Theorem 1 is an immediate consequence of Theorem 1a in view of Lemmas 1.1 and 1.2.

4. **The difference equation with one linear coefficient.** We now deal with equation (1.5), and as in the previous case, we obtain a slightly more general result expressed in terms of an integral equation(7).

THEOREM 2a. *Let $p(\lambda)$ be a right-continuous function of bounded variation on every finite interval and such that*

$$N_1(p) \equiv \int_{-\infty}^{0} e^{\alpha\lambda}\,|\,dp(\lambda)\,| + \int_{0}^{\infty} e^{\beta\lambda}\,|\,dp(\lambda)\,| < \infty. \tag{4.1}$$

Let $y_0 = \operatorname{Im}[p(0+0)-p(0-0)]$ and let $a<\operatorname{Im} z<b$ be any strip whose closure does not contain the line $\operatorname{Im} z = y_0$. Let $g(z)$ be a function analytic in the strip $a<\operatorname{Im} z<b$ and of finite norm:

$$\bar{N}(g) = \sup_{a<y<b}\left[\int_{-\infty}^{0} e^{\beta x}\,|\,g(x+iy)\,|\,dx + \int_{0}^{\infty} e^{\alpha x}\,|\,g(x+iy)\,|\,dx < \infty. \right. \tag{4.2}$$

Then the equation

$$zf(z) = \int_{-\infty}^{\infty} f(z-\lambda)dp(\lambda) + g(z) \tag{4.3}$$

has one and only one solution $f(z)$ which is analytic in the strip $a<\operatorname{Im} z<b$ and which is of finite norm, $N(f)<\infty$.

We base the proof upon the following lemma.

LEMMA 4.1. *Let $p(\lambda)$ be a function of bounded variation on every finite interval and of finite norm, $N_1(p)<\infty$. Let $z_1 = x_1+iy_1$ be any complex number, and let $a<\operatorname{Im} z<b$ be any strip whose distance from the line $\operatorname{Im} z = y_1$ is greater than $N_1(p)$. Let $g(z)$ be a function analytic in the strip $a<\operatorname{Im} z<b$ for which $N(g)<\infty$. Then the equation*

(7) This result and its proof were suggested by applying certain duality considerations to a corresponding result on differential equations. See R. H. Cameron, *Quadratures involving trigonometric sums*, Journal of Mathematics and Physics vol. 19 (1940) pp. 161–166. This result is however certainly not a direct dual since $g(z)$ need not have any Fourier transform.

$$(4.4) \qquad (z - z_1)f(z) = \int_{-\infty}^{\infty} f(z - \lambda)dp(\lambda) + g(z)$$

has one and only one solution $f(z)$ of finite norm and analytic in the strip $a<\operatorname{Im} z<b$. In addition

$$(4.5) \qquad N[f; y] \leqq \frac{N(g; y)}{|y - y_1| - N_1(p)}, \qquad a < y < b.$$

We shall prove Lemma 4.1 by the method of successive approximations. Let

$$f_0(z) = g(z)/(z - z_1)$$

and

$$(4.6) \qquad f_n(z) = \frac{1}{z - z_1}\int_{-\infty}^{\infty} f_{n-1}(z - \lambda)dp(\lambda), \qquad n = 1, 2, \cdots,$$

each $f_n(z)$ existing in view of Lemma 2.1. We shall show that

$$(4.7) \qquad f(z) = \sum_{n=0}^{\infty} f_n(z)$$

also exists and satisfies the conclusions of Lemma 4.1. First, by Lemma 2.1,

$$N(f_0; y) \leqq \frac{1}{|y - y_1|} N(g; y)$$

and

$$N(f_n; y) \leqq \frac{1}{|y - y_1|} N(f_{n-1}; y)N_1(p) \leqq \frac{[N_1(p)]^n}{|y - y_1|^{n+1}} N(g; y),$$
$$a < g < b, n = 1, 2, \cdots.$$

Thus

$$(4.8) \qquad \sum_{n=0}^{\infty} N(f_n; y) \leqq \sum_{n=0}^{\infty} \frac{[N_1(p)]^n}{|y - y_1|^{n+1}} N_1(g; y) = \frac{N(g; y)}{|y - y_1| - N_1(p)}.$$

By Lemmas 2.4, 2.2, and equation (4.8) we see that the function defined by (4.7) is analytic and of finite norm in the strip $a<\operatorname{Im} z<b$ and that (4.5) holds.

To see that the f defined by (4.7) satisfies equation (4.4) we note by (4.6) that

$$(4.9) \qquad f(z) = f_0(z) + \sum_{n=1}^{\infty} f_n(z) = \frac{1}{z - z_1}\left[g(z) + \sum_{n=1}^{\infty}\int_{-\infty}^{\infty} f_{n-1}(z - \lambda)dp(\lambda)\right].$$

By Lemma 2.4, we can interchange the order of summation and integration and thus

$$f(z) = \frac{1}{z - z_1}\left[g(z) + \int_{-\infty}^{\infty} f(z - \lambda)dp(\lambda)\right], \tag{4.10}$$

that is, $f(z)$ satisfies equation (4.4).

For uniqueness let us assume that $f^*(z)$ is a solution of (4.4), analytic in $a < \text{Im } z < b$ and of finite norm. Then

$$f^*(z) - f(z) = \frac{1}{z - z_1}\int_{-\infty}^{\infty} [f^*(z - \lambda) - f(z - \lambda)]dp(\lambda)$$

and

$$N(f^* - f; y) \leqq \frac{1}{|y - y_1|} N(f^* - f; y)N_1(p), \qquad a < y < b,$$

or

$$N(f^* - f; y)\left(1 - \frac{N_1(p)}{|y - y_1|}\right) \leqq 0, \qquad a < y < b.$$

Since the second factor is positive, $f^* - f \equiv 0$.

This concludes the proof of Lemma 4.1.

We now proceed to the proof of Theorem 2a itself. Let us therefore return to the notation of that theorem and assume that the hypotheses of the theorem are satisfied. We begin by introducing a transformation which transforms equation (4.3) into one to which the preceding lemma applies. This transformation will be based upon the following function:

$$R(w) = \exp\left[\int_{-\infty}^{-\eta} \frac{e^{\lambda w}}{\lambda} dp(\lambda) + \int_{\eta}^{\infty} \frac{e^{\lambda w}}{\lambda} dp(\lambda)\right], \qquad \alpha < \text{Re } w < \beta, \tag{4.11}$$

the η being chosen positive and so small that

$$N_1(p_1) < \min\{|y_0 - a|, |y_0 - b|\}, \tag{4.12}$$

where

$$p_1(\lambda) = \begin{cases} p(\eta) - p(0 + 0), & \lambda \geqq \eta, \\ p(\lambda) - p(0 + 0), & 0 \leqq \lambda \leqq \eta, \\ p(\lambda) - p(0 - 0), & -\eta \leqq \lambda < 0, \\ p(-\eta) - p(0 - 0), & \lambda \leqq -\eta. \end{cases} \tag{4.13}$$

This is possible since by hypothesis y_0 lies outside the closed interval $a \leqq y \leqq b$ so the right member of (4.12) is positive, and the variation of $p_1(\lambda)$ can be made arbitrarily small due to the continuity of $p_1(\lambda)$ at the origin. Then if

$$j(\lambda) = \begin{cases} 0, & \lambda \geqq 0, \\ -1, & \lambda < 0, \end{cases}$$

and([8]) $z_0=x_0+iy_0=p(0+0)-p(0-0)$, we have

$$R(w) = \exp\left\{\int_{-\infty}^{\infty} \frac{e^{\lambda w}}{\lambda} d[p(\lambda) - p_1(\lambda) - z_0 j(\lambda)]\right\}. \tag{4.14}$$

We next show that there exist functions $r(\lambda)$ and $r^*(\lambda)$ such that

$$R(w) = \int_{-\infty}^{\infty} e^{\lambda w} dr(\lambda) \tag{4.15}$$

and

$$\frac{1}{R(w)} = \int_{-\infty}^{\infty} e^{\lambda w} dr^*(\lambda), \tag{4.16}$$

and

$$j(\lambda) = \int_{-\infty}^{\infty} [r(\lambda - t) - r(-t)] dr^*(t). \tag{4.17}$$

For let $q_0(\lambda)=j(\lambda)$ and

$$q_1(\lambda) = \int_0^{\lambda+0} \frac{d[p(\xi) - p_1(\xi) - z_0 j(\xi)]}{\xi} \tag{4.18}$$

and

$$q_n(\lambda) = \int_{-\infty}^{\infty} [q_1(\lambda - \mu) - q_1(-\mu)] dq_{n-1}(\mu), \qquad n = 2, 3, \cdots. \tag{4.19}$$

Since q_0 and q_1 are normalized it follows by the corollary to Lemma 2.6 that all the q_n are normalized.

Now

$$\int_{-\infty}^{\infty} e^{\lambda w} dq_n(\lambda) = \left[\int_{-\infty}^{\infty} e^{\lambda w} dq_1(\lambda)\right]^n, \qquad n = 0, 1, 2, \cdots.$$

Next define

$$r(\lambda) = \sum_{n=0}^{\infty} \frac{1}{n!} q_n(\lambda) \tag{4.20}$$

and

$$r^*(\lambda) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} q_n(\lambda). \tag{4.21}$$

([8]) Note that this is consistent with the definition of y_0 given in the statement of Theorem 2a.

By inspection $q_0(\lambda)$ is of finite norm, and the same is true of $q_1(\lambda)$, when we note that the function with respect to which we integrate is constant in a neighborhood of the origin. By equation (2.18) it follows that each q_n is of finite norm with $N_1(q_n) \leqq [N_1(q_1)]^n$, and by Lemma 2.7 that $r(\lambda)$ and $r^*(\lambda)$ defined in (4.20) and (4.21) exist and are of finite norm. Then

$$R(w) = \exp\left\{\int_{-\infty}^{\infty} \frac{e^{\lambda w}}{\lambda} d[p(\lambda) - p_1(\lambda) - z_0 j(\lambda)]\right\}$$

$$= \exp\left\{\int_{-\infty}^{\infty} e^{\lambda w} dq_1(\lambda)\right\}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!}\left[\int_{-\infty}^{\infty} e^{\lambda w} dq_1(\lambda)\right]^n = \sum_{n=0}^{\infty} \frac{1}{n!}\int_{-\infty}^{\infty} e^{\lambda w} dq_n(\lambda).$$

We shall show that the integration and summation can be interchanged, thus giving

$$R(w) = \int_{-\infty}^{\infty} e^{\lambda w} d\left[\sum_{n=0}^{\infty} \frac{1}{n!} q_n(\lambda)\right] = \int_{-\infty}^{\infty} e^{\lambda w} dr(\lambda).$$

The interchange is justified by the following argument:

$$\sum_{n=0}^{\infty} \frac{1}{n!}\int_{-M}^{M} e^{\lambda w} dq_n(\lambda)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!}\left[e^{wM} q_n(M) - e^{-wM} q_n(-M) - w\int_{-M}^{M} q_n(\lambda) e^{\lambda w} d\lambda\right]$$

$$= e^{wM} r(M) - e^{-wM} r(-M) - w\int_{-M}^{M} r(\lambda) e^{\lambda w} d\lambda = \int_{-M}^{M} e^{\lambda w} dr(\lambda).$$

By the Weierstrass M-test and the convergence of

$$\sum_{n=1}^{\infty} \frac{1}{n!} N_1(q_n)$$

it follows that the above series in the first member converges uniformly for all real M and thus as $M \to \infty$ the limit of the sum is the sum of the limits. Thus (4.15) holds; and it is obvious that (4.16) can be proved in the same way. Moreover, (4.15) and (4.16) imply (4.17). Returning now to (4.14), we differentiate and obtain

$$dR(w)/dw = R(w)[P(w) - P_1(w) - z_0] \tag{4.22}$$

where

$$P(w) = \int_{-\infty}^{\infty} e^{\lambda w} dp(\lambda), \qquad P_1(w) = \int_{-\infty}^{\infty} e^{\lambda w} dp_1(\lambda).$$

Rewriting (4.22) by replacing each side by the function of which it is a Fourier-Stieltjes transform, and noting that

$$\frac{dR(w)}{dw} = \int_{-\infty}^{\infty} e^{\lambda w} d_\lambda \int_0^\lambda \mu dr(\mu)$$

we obtain

$$(4.23) \qquad \int_0^\lambda \mu dr(\mu) = \int_{-\infty}^{\infty} [r(\lambda - \mu) - r(-\mu)] d[p(\mu) - p_1(\mu) - z_0 j(\mu)].$$

Now consider the functional equation

$$(4.24) \qquad (z - z_0) f_1(z) = \int_{-\infty}^{\infty} f_1(z - \lambda) dp_1(\lambda) + g_1(z)$$

where

$$(4.25) \qquad g_1(z) = \int_{-\infty}^{\infty} g(z - \lambda) dr^*(\lambda).$$

It is clear by (4.12) that the functional equation (4.24) satisfies the hypothesis of Lemma 4.1, and thus has one and only one analytic solution $f_1(z)$ of finite norm in $a < \text{Im } z < b$. We can therefore consider that $f_1(z)$ is completely defined by (4.24), and in terms of it we define

$$(4.26) \qquad f(z) = \int_{-\infty}^{\infty} f_1(z - t) dr(t), \qquad a < \text{Im } z < b.$$

We shall show that this $f(z)$ is the unique analytic solution of finite norm of the original equation (4.3). By Lemmas 2.3 and 2.1, $f(z)$ so defined is analytic and of finite norm. Also by (4.24) we have

$$\int_{-\infty}^{\infty} (z - t - z_0) f_1(z - t) dr(t) = \int_{-\infty}^{\infty} dr(t) \int_{-\infty}^{\infty} f_1(z - \lambda - t) dp_1(\lambda) + \int_{-\infty}^{\infty} g_1(z - t) dr(t).$$

Transposing, simplifying, and using (4.25) and (4.17), we find

$$z \int_{-\infty}^{\infty} f_1(z - t) dr(t) = \int_{-\infty}^{\infty} f_1(z - t) t dr(t) + z_0 \int_{-\infty}^{\infty} f_1(z - t) dr(t) + \int_{-\infty}^{\infty} f_1(z - t) d_t \int_{-\infty}^{\infty} [r(t - \lambda) - r(-\lambda)] dp_1(\lambda) + g(z).$$

By (4.26) and (4.23), this yields

$$zf(z) = \int_{-\infty}^{\infty} f_1(z-t)d_t \int_{-\infty}^{\infty} [r(t-\lambda) - r(-\lambda)]d[p(\lambda) - z_0 j(\lambda)] + z_0 f(z) + g(z)$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f_1(z-\mu-\lambda)dr(\mu) \right] dp(\lambda) + g(z)$$

$$= \int_{-\infty}^{\infty} f(z-\lambda)dp(\lambda) + g(z).$$

Thus $f(z)$, defined in (4.26), satisfies the original equation (4.3).

Similarly if any $f(z)$ satisfies (4.3), $f(z)$ assumed analytic and of finite norm in $a < \text{Im } z < b$, it is seen that the function

$$f_1(z) = \int_{-\infty}^{\infty} f(z-\lambda)dr^*(\lambda)$$

satisfies equation (4.24). Thus the uniqueness of the solution of (4.24) implies the uniqueness of the solution of (4.3).

This concludes the proof of Theorem 2a.

Theorem 2 follows from Theorem 2a and Lemmas 1.1 and 1.2.

5. The difference equation with linear coefficients.

THEOREM 3a. *Let $p(\lambda)$ and $q(\lambda)$ be complex-valued functions of bounded variation on every finite interval such that*

$$N_1(p) < \infty, \quad N_1(q) < \infty, \quad \text{singular part of } p = 0. \tag{5.1}$$

Moreover let the analytic function

$$P(w) = \int_{-\infty}^{\infty} e^{\lambda w} dp(\lambda) \tag{5.2}$$

be non-vanishing in the strip $\alpha < \text{Re } w < \beta$. Define(9)

$$y_0 = \text{Im}\left[\lim_{T\to\infty} \frac{1}{Ti} \int_{\gamma}^{\gamma+Ti} \frac{\int_{-\infty}^{\infty} e^{\lambda w} dq(\lambda)}{\int_{-\infty}^{\infty} e^{\lambda w} dp(\lambda)} dw \right]. \tag{5.3}$$

Let $a < \text{Im } z < b$ be any strip whose closure does not contain the line $\text{Im } z = y_0$. Let $g(z)$ be a function analytic in the strip $a < \text{Im } z < b$ and of finite norm,

$$N(g) < \infty. \tag{5.4}$$

Then the equation

$$\int_{-\infty}^{\infty} (z-\lambda)f(z-\lambda)dp(\lambda) - \int_{-\infty}^{\infty} f(z-\lambda)dq(\lambda) = g(z) \tag{5.5}$$

has one and only one analytic solution $f(z)$ of finite norm in the strip $a < \text{Im } z < b$.

(9) This limit exists and is independent of γ for $\alpha < \gamma < \beta$.

For the proof we again use Pitt's theorem (loc. cit.) which yields the result that $1/P(w)$ is expressible in the form (3.5) with $N_1(p^*) < \infty$.

We next convolve equation (5.5) with $p^*(\mu)$ and obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(z-\lambda-\mu)f(z-\lambda-\mu)dp(\lambda)dp^*(\mu) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(z-\lambda-\mu)dq(\lambda)dp^*(\mu) = \int_{-\infty}^{\infty}g(z-\mu)dp^*(\mu),$$

or

$$zf(z) - \int_{-\infty}^{\infty}f(z-\lambda)d\tilde{q}(\lambda) = \tilde{g}(z) \tag{5.6}$$

where

$$\tilde{q}(\lambda) = \int_{-\infty}^{\infty}[q(\lambda-\mu) - q(-\mu)]dp^*(\mu), \quad \tilde{g}(z) = \int_{-\infty}^{\infty}g(z-\mu)dp^*(\mu). \tag{5.7}$$

We note by Lemmas 2.3, 2.6, and 2.1 that $\tilde{g}(z)$ is analytic in $a < \text{Im } z < b$ and that

$$N_1(\tilde{q}) < \infty, \qquad N(\tilde{g}) < \infty. \tag{5.8}$$

Moreover[10]

$$y_0 = \text{Im } [\tilde{q}(0+0) - \tilde{q}(0-0)]. \tag{5.9}$$

Now we apply Theorem 2a to (5.6). We note that the hypotheses of that theorem are satisfied with p and g replaced by $\tilde{q}$, $\tilde{g}$. Thus there is a unique analytic solution $f(z)$ of (5.6) of finite norm. Hence if (5.5) has an analytic solution of finite norm it is unique. But the solution of (5.6) obviously satisfies (5.5) as we see by convolving equation (5.6) with p. This yields the theorem.

As in the earlier cases it is easily seen that Theorem 3 follows from Theorem 3a.

---

[10] That (5.9) agrees with (5.3) is easily seen.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

A GENERALIZED THEORY OF DYNAMICAL TRAJECTORIES

BY

EDWARD KASNER AND JOHN DeCICCO

1. Introduction. The differential geometry of the dynamical trajectories of positional fields of force has been developed in the Princeton Colloquium(1). The fields of force, considered heretofore, depend only upon the position of the point. We shall call a field of force *generalized* if it depends not only upon the position of the point but also upon the direction through the point. In this paper, it is our purpose to begin the development of the geometry of the dynamical trajectories of such generalized fields of force in the plane. Positional fields may be described as isotropic and generalized fields as *anisotropic*.

In a generalized field of force, there are ∞^3 dynamical trajectories. These are represented by a differential equation of the form

$$y''' = G(x, y, y')y'' + H(x, y, y')y''^2. \tag{G}$$

Conversely, any system of curves of the type (G) may represent the dynamical trajectories of $\infty^{f(2)}$ generalized fields of force(2). We therefore find it appropriate to term any system of curves of the type (G) as *generalized dynamical trajectories*.

Of course, not every such system of curves represents the trajectories of a positional field of force. Kasner proved that the family of ordinary dynamical trajectories is characterized by a set of *five* independent geometric properties. Of these, the Property I is equivalent analytically to stating that a system of ordinary dynamical trajectories must be represented by a differential equation of the type (G). The other four geometric properties specialize the functions G and H.

The Property I is described as follows. For each of the ∞^1 curves of the ∞^3 trajectories which pass through a given lineal element E, construct the osculating parabola at E. The Property I is given by any one of the following three equivalent statements(3).

Presented to the Society, October 31, 1942; received by the editors August 28, 1942.

(1) Kasner, *Differential-geometric aspects of dynamics*, Amer. Math. Soc. Colloquium Publications, vol. 3, 1913, 1934 (referred to as Princeton Colloquium). Also see a series of papers in Trans. Amer. Math. Soc. vols. 7–11 (1906–1910).

(2) The symbol $\infty^{f(2)}$ means that the force vector contains an arbitrary function of two variables. See Kasner, *A notation for infinite manifolds*, Amer. Math. Monthly vol. 49 (1942) pp. 243–244.

(3) Recently Terracini has given an alternate projective characterization of Property I by conic sections. See *Sobre la ecuacion diferencial* $y'''=G(x, y, y')y''+H(x, y, y')y'^2$, Revista de Matematicas vol. 2 (1941) pp. 245–329.

(I_a) The locus of the foci is a circle through the point of E.

(I_b) The directrices form a pencil of lines.

(I_c) The envelope of the ∞^1 parabolas is a straight line.

Next we shall compare the ∞^1 lines of force and the ∞^2 rest trajectories. Let a particle which is initially at rest start from a point O. The rest trajectory will then be tangent to the line of force at O. We study the ratio ρ of the departure of the rest trajectory to that of the line of force at O. In the positional case, Kasner showed that $\rho=1/3$, or, if the order of contact of each of the curves with their common tangent line is n, $\rho=1/(2n+1)$. We obtain a converse theorem determining all generalized fields of force for which this ratio assumes the above values. (See formulas (35) and (45).)

Finally we shall define for our generalized fields of force other systems of curves, namely, velocity systems, systems S_k, and pressure curves. Each of these systems is of the type (G) and hence possesses the Property I. While the systems S_k and the pressure curves may be represented by a differential equation of the type (G), it is found that velocity systems must be curvature trajectories(4).

2. **The differential equation of the generalized dynamical trajectories.** Let ϕ and ψ be the rectangular components parallel to the x- and y-axes of the generalized force vector acting at the lineal element (x, y, y')(5). We assume that our force vector (ϕ, ψ) does not identically agree with the direction of the corresponding lineal element E. For, if $\psi-y'\phi\equiv 0$, the trajectories are straight lines, and are, therefore, of no interest. The equations of motion of a particle of unit mass are

$$d^2x/dt^2 = \phi(x, y, y'), \qquad d^2y/dt^2 = \psi(x, y, y'). \tag{1}$$

The particle may be started from any position (x_0, y_0) with any velocity $(dx_0/dt_0, dy_0/dt_0)$. A definite trajectory is described. Since the same curve may be obtained by starting from any one of its ∞^1 points, the total number of trajectories, for all initial conditions, is ∞^3.

We now proceed to eliminate the time t from the equations (1). (For the positional case, see the Princeton Colloquium, pp. 7–9.) The elimination can be performed directly from (1), but we prefer to do it in the following manner. Let N and T be the normal and tangential components along a trajectory. Then

$$\begin{aligned} N &= (\psi - y'\phi)/(1 + y'^2)^{1/2}, \qquad T = (\phi + y'\psi)/(1 + y'^2)^{1/2}; \\ \phi &= (T - y'N)/(1 + y'^2)^{1/2}, \qquad \psi = (N + y'T)/(1 + y'^2)^{1/2}. \end{aligned} \tag{2}$$

(4) Kasner, *Dynamical trajectories and curvature trajectories*, Bull. Amer. Math. Soc. vol. 40 (1934) pp. 449–455.

(5) The functions ϕ and ψ could have been assumed to involve x and y and also to be homogeneous of degree zero in dx/dt and dy/dt. Then since during motion dx/dt and dy/dt are not both zero, the functions ϕ and ψ are seen to depend on (x, y, y') only.

By these equations, it is observed that either ϕ and ψ, or $N \not\equiv 0$ and T, are two arbitrary functions of (x, y, y').

If v denotes the speed of the particle and r the radius of curvature, we find

$$v^2 = rN, \qquad vv_s = T. \tag{3}$$

Eliminating the speed v from these equations, we find the intrinsic differential equation of the generalized dynamical trajectories to be

$$d(rN)/ds = 2T. \tag{4}$$

Upon putting this equation into the cartesian form, we find the following result.

THEOREM 1. *The ∞^3 dynamical trajectories of a generalized field of force are given by the differential equation*

$$(\psi - y'\phi)y''' = [\psi_x + y'(\psi_y - \phi_x) - y'^2\phi_y]y'' + [\psi_{y'} - y'\phi_{y'} - 3\phi]y''^2. \tag{5}$$

This is of type (G).

It is noted that this represents the ordinary dynamical type if ϕ and ψ are independent of y'.

3. **The type** (G) **always represents the trajectories of a generalized field of force.** In order to simplify our discussion for generalized fields of force, it is convenient to introduce the functions α and β defined as follows:

$$\begin{aligned} e^\alpha &= (1 + y'^2)^{3/2}N, & \beta &= -2T/(1 + y'^2)N; \\ N &= e^\alpha/(1 + y'^2)^{3/2}, & T &= -e^\alpha\beta/2(1 + y'^2)^{1/2}. \end{aligned} \tag{6}$$

We shall also write the equations connecting (ϕ, ψ) and (α, β) as follows:

$$\begin{aligned} e^\alpha &= (1 + y'^2)(\psi - y'\phi), & \beta &= -2(\phi + y'\psi), \\ \phi &= -\frac{[2y'e^\alpha + (1 + y'^2)\beta]}{2(1 + y'^2)^2}, & \psi &= \frac{[2e^\alpha - y'(1 + y'^2)\beta]}{2(1 + y'^2)^2}. \end{aligned} \tag{7}$$

Now placing the last two of equations (6) into the differential equation (5), we see that the differential equation of the generalized dynamical trajectories is

$$y''' = (\alpha_x + y'\alpha_y)y'' + (\alpha_{y'} + \beta)y''^2. \tag{8}$$

THEOREM 2. *Any system of curves of the type* (G) *represents the dynamical trajectories of $\infty^{f(2)}$ generalized fields of force.*

We proceed to the proof of our theorem. Let a system of curves of the type (G) be given. If there exists a generalized field of force of which the given curves are the trajectories, the functions α and β must be determined so that

$$\alpha_x + y'\alpha_y = G, \qquad \alpha_{y'} + \beta = H. \tag{9}$$

By the theory of differential equations, we know that the first equation determines α as an expression of the form

$$\alpha = \alpha_1(x, y, y') + \alpha_2(y - xy', y'), \tag{10}$$

where α_1 is a definite function of three variables and α_2 is an arbitrary function of two variables. The second equation then completely determines β. This completes the proof of our Theorem 2.

For example, all generalized fields of force whose dynamical trajectories are the vertical parabolas $y=ax^2+bx+c$, are determined by

$$\alpha = F(y - xy', y'), \qquad \beta = xF_{y-xy'} - F_{y'}. \tag{11}$$

4. Fields of force depending only upon differential elements of order n for $n \geqq 2$. The argument presented in the preceding section may be generalized to prove the following result.

THEOREM 3. *Any system of ∞^{n+1} curves for $n \geqq 2$ may represent the dynamical trajectories of $\infty^{f(n+2)}$ more extensive fields of force depending only upon differential elements of order n.*

The equations of motion of a particle of unit mass for a field of force which depends only upon the position of a differential element of order n, are

$$d^2x/dt^2 = \phi(x, y, y', \cdots, y^{(n)}), \qquad d^2y/dt^2 = \psi(x, y, y', \cdots, y^{(n)}). \tag{12}$$

Now if N and T represent the normal and tangential components of the force vector along any trajectory, it is found by equations analogous to (2) that N and T may be considered to be any two arbitrary functions of $(x, y, y', \cdots, y^{(n)})$. Then since $n \geqq 2$, the functions N^* and T^* defined by

$$N^* = rN, \qquad T^* = 2T(1 + y'^2)^{1/2}, \tag{13}$$

are also arbitrary functions of $(x, y, y', \cdots, y^{(n)})$.

The intrinsic differential equation of the trajectories of such an extensive field of force is given by an equation analogous to (4). Therefore replacing rN and $2T$ by the functions N^* and $T^*(1+y'^2)^{-1/2}$ and performing the differentiation, the differential equation of the dynamical trajectories is

$$y^{(n+1)} = (1/N^*_{y^{(n)}})\left[T^* - (N^*_x + y'N^*_y + y''N^*_{y'} + \cdots + y^{(n)}N^*_{y^{(n-1)}})\right]. \tag{14}$$

Therefore there are ∞^{n+1} dynamical trajectories in an extensive field of force depending only upon the position of a differential element of order $n \geqq 2$.

The right-hand side of (14) may represent any arbitrary function of $(x, y, y', \cdots, y^{(n)})$. Also the choice of T^* is uniquely determined as soon as

N^* is given. Therefore there are $\infty^{f(n+2)}$ extensive fields of force whose trajectories are a given family of ∞^{n+1} curves.

5. **The geometric Property I.** Before discussing the Property I, it is advisable to collect here the various formulas for the osculating parabola of a curve(6). The equation in running coordinates (X, Y) of the osculating parabola at any differential element of the third order (x, y, y', y'', y''') of the curve $y=y(x)$ is

$$[(y'y''' - 3y''^2)(X - x) - y'''(Y - y)]^2 + 18y''^3[y'(X - x) - (Y - y)] = 0. \tag{15}$$

The focus is

$$X + iY = x + iy + \frac{3y''(1 + iy')^2}{2[y'''(1 + iy') - 3iy''^2]}. \tag{16}$$

The equation of the directrix is

$$2y'''(X - x) + 2(y'y''' - 3y''^2)(Y - y) = 3y''(1 + y'^2). \tag{17}$$

Consider now any system of ∞^3 curves. For each of the ∞^1 curves of this system which pass through a given lineal element E, construct the osculating parabola at E. This gives rise to ∞^1 parabolas all containing the lineal element E.

THEOREM 4. *A system of ∞^3 curves is a generalized dynamical system if and only if the locus of the foci of the ∞^1 osculating parabolas, constructed at any lineal element E, is a circle passing through the point of E.*

The equation of this focal circle is

$$2G[(X - x)^2 + (Y - y)^2] - [y'(1 + y'^2)H + 3(1 - y'^2)](X - x) + [(1 + y'^2)H - 6y'](Y - y) = 0. \tag{18}$$

THEOREM 5. *A system of ∞^3 curves is a generalized dynamical system if and only if the directrices of the ∞^1 osculating parabolas, constructed at any lineal element E, form a pencil of lines.*

The vertex D of this pencil of lines is

$$D: \quad X + iY = x + iy + \frac{(1 + y'^2)}{2G}[-(y'H - 3) + iH]. \tag{19}$$

Consider now any system of ∞^3 curves. At a fixed lineal element E, y''' is a function of y'' only. Now let ω represent the expression y'''/y''^2 so that at E,

(6) For an analogous discussion in minimal coordinates, see Kasner and DeCicco, *Families of curves conformally equivalent to circles*, Trans. Amer. Math. Soc. vol. 49 (1941) pp. 378–391.

ω is a function of y'' only. By (15) the ∞^1 osculating parabolas at E are represented by the equation

$$(20) \qquad y''[(y'\omega - 3)(X - x) - \omega(Y - y)]^2 + 18[y'(X - x) - (Y - y)] = 0.$$

Now if $\omega_{y''}=0$, our family of parabolas is linear. The foci will all lie on the straight line into which the focal circle (18) degenerates by letting G approach zero. The directrices will all be parallel, their directions being perpendicular to the direction so that the line of E bisects the angle at the point O of E formed by this direction and the focal straight line. The complete envelope of this linear family of parabolas consists of the line of E and the line at infinity.

Any system of ∞^3 curves such that the ∞^1 osculating parabolas, constructed at E, form a linear family must be of the special subclass of type (G)

$$(21) \qquad y''' = H(x, y, y')y''^2.$$

Let now $\omega_{y''}\neq 0$. Thence our family of ∞^1 parabolas (20) possesses an envelope (in addition to the common tangent line of the element E). This envelope is given parametrically by the equations

$$(22) \qquad X - x = -\frac{3(\omega + 2y''\omega_{y''})}{2y''^3\omega_{y''}^2}, \quad (Y - y) - y'(X - x) = \frac{9}{2y''^3\omega_{y''}^2}.$$

Now differentiating these partially with respect to y'', we find

$$(23) \qquad \begin{aligned} X_{y''} &= \frac{3}{2y''^4\omega_{y''}^3}(3\omega_{y''} + 2y''\omega_{y''y''})(\omega + y''\omega_{y''}), \\ Y_{y''} - y'X_{y''} &= -\frac{9}{2y''^4\omega_{y''}^3}(3\omega_{y''} + 2y''\omega_{y''y''}). \end{aligned}$$

Our envelope is obviously a point if and only if $3\omega_{y''}+2y''\omega_{y''y''}=0$. Therefore the family of ∞^1 osculating parabolas, constructed at the element E, pass through a fixed point if and only if the differential equation of the system of ∞^3 curves is of the form

$$(24) \qquad y''' = \mu(x, y, y')y''^{3/2} + \nu(x, y, y')y''^2.$$

Now we exclude this situation. Then the slope dY/dX of our envelope (21) is given by the equation

$$(25) \qquad dY/dX - y' = -3/(\omega + y''\omega_{y''}).$$

This formula demonstrates that our envelope is a straight line if and only if $y''\omega=G+Hy''$. Therefore we may state the following result.

THEOREM 6. *A system of ∞^3 curves is a generalized dynamical system if and*

only if the envelope of the ∞^1 osculating parabolas, constructed at any lineal element E, is a straight line (not passing through the point of E).

This common tangent line γ of the ∞^1 osculating parabolas is given by the equation

$$2G(y'H - 3)(X - x) - 2GH(Y - y) + 9 = 0. \tag{26}$$

Thus our ∞^1 parabolas are tangent to the line γ, the line of the element E at the point of E, and the line at infinity.

Our line γ intersects the line of the lineal element E in the point

$$P: \quad X + iY = x + iy + (3/2G)(1 + iy'). \tag{27}$$

Moreover we find that our line γ is tangent to the focal circle at this point P. The line DP connecting the vertex D of the directrices and this point P, is perpendicular to the line of E.

We remark that the one-parameter set of parabolas for a given element E, is in general quadratic.

6. **The angular rate λ and the terminal curve C.** For the further development of our theory, it is found essential in a given generalized field of force to associate a number λ to any lineal element E. As the element E rotates about its point O, the corresponding force vector F also rotates about O. We define the *angular rate* λ as the instantaneous rate of change of the inclination of F with respect to the inclination of E. It is given by the formula

$$\lambda = \frac{1 + y'^2}{\phi^2 + \psi^2}(\phi\psi_{y'} - \psi\phi_{y'}). \tag{28}$$

Also as E rotates about its point O, the end point of the corresponding force vector F describes a curve C. We shall call this the *terminal curve* C corresponding to the point O. The terminal curve C degenerates into a fixed point if and only if our field of force is ordinary positional.

Obviously for an ordinary positional field of force the angular rate $\lambda \equiv 0$. However there are also certain generalized fields of force for which the angular rate $\lambda \equiv 0$. These fields are such that the terminal curve corresponding to any point O is a straight line through O.

7. **The lines of force and the rest trajectories.** In this section, we wish to define two special important families of curves in a given generalized field of force, namely, the lines of force and the rest trajectories. Let us first consider the lines of force. In general, at a given point O, there is one lineal element E_0 such that the direction of the force vector F_0 is identical with that of E_0. Let us also suppose that the angular rate λ at such an element E_0 is not unity. If in a certain region of the plane, there are ∞^2 such lineal elements E_0, we define the integral curves of this set of ∞^2 elements E_0 as the *lines of force* of our generalized field of force.

The lines of force are given by the differential equation of the first order

$$\psi(x, y, y') - y'\phi(x, y, y') = 0. \tag{29}$$

The condition that the angular rate $\lambda \neq 1$ is equivalent to stating that there exists a function $y' = f(x, y)$ which satisfies this equation identically. Thus in general, there are ∞^1 lines of force.

A *rest trajectory* is defined as the path described in a generalized field of force by a particle which starts at rest from a given point O. From the equations

$$dy/dt = y'\,dx/dt, \qquad d^2y/dt^2 = y''(dx/dt)^2 + y'\,d^2x/dt^2, \tag{30}$$

together with the initial conditions $dx_0/dt_0 = 0$, $dy_0/dt_0 = 0$, it is immediately deduced that a rest trajectory is always initially tangent to the line of force through the given point O. Since on any rest trajectory, there is, in general, only one such point O where the particle is at rest, it is deduced that there are ∞^2 rest trajectories in a given generalized field of force(7).

8. **The generalization of the theorem concerning one-third of the curvatures.**(8) We shall now study in a given generalized field of force the ratio ρ of the departure of the rest trajectory to that of the line of force at a given point O. For the ordinary case Kasner showed that this ratio $\rho = 1/3$. We shall find, however, a more general formula for ρ in terms of the angular rate λ.

At the point O, let E_0 be the lineal element of the line of force. Then the rest trajectory described by a particle initially at rest at O will be tangent to the line of E_0.

Now we shall assume that the curvature of the rest trajectory is not zero at the point O. Let us calculate the value of the second derivative y'' of the rest trajectory at O. Since our rest trajectory is tangent to the line of force at O, the first derivative y' must satisfy the equation (29). Therefore from this and the equation (5), it results that the second derivative y'' of the rest trajectory at O is given by the formula

$$[\psi_x + y'(\psi_y - \phi_x) - y'^2\phi_y] + [\psi_{y'} - y'\phi_{y'} - 3\phi]y'' = 0. \tag{31}$$

---

(7) Of course, we are assuming that in the region of the *opulence* (the totality of ∞^3 lineal elements of the plane) where the generalized field of force is defined, there are points O (even in the imaginary domain) through which there pass such special lineal elements E_0 (one or more). Otherwise there are no theories of lines of force and rest trajectories.

(8) For a generalization to acceleration fields of higher order, see Kasner and Mittleman, *A general theorem on the initial curvature of dynamical trajectories*, Proc. Nat. Acad, Sci. U.S.A. vol. 28 (1942) pp. 48–52; see also *Extended theorems in dynamics*, Science vol. 95 (1942) pp. 249–250.

Also see Fialkow, *Initial motion in fields of force*, Trans. Amer. Math. Soc. vol. 40 (1936) pp. 495–501; and Kasner and Fialkow, Trans. Amer. Math. Soc. vol. 41 (1937).

The second derivative Y'' of the line of force is calculated by taking the total derivative of the equation (29). Therefore this is given by

$$[\psi_x + y'(\psi_y - \phi_x) - y'^2\phi_y] + [\psi_{y'} - y'\phi_{y'} - \phi]Y'' = 0. \tag{32}$$

Subtracting these two equations, we find

$$[\psi_{y'} - y'\phi_{y'} - 3\phi]y'' = [\psi_{y'} - y'\phi_{y'} - \phi]Y''. \tag{33}$$

Since the magnitude of the force by assumption is not zero and since at E_0, $\psi = y'\phi$, we see that $\phi \neq 0$. Upon dividing this equation by ϕ and making use of the equation (28) defining our angular rate λ, we see that the preceding equation may be written in the form

$$(3 - \lambda)y'' = (1 - \lambda)Y''. \tag{34}$$

The angular rate $\lambda \neq 1$. For, otherwise, the line of force could not exist. Also if the angular rate $\lambda \neq 3$, the curvature of the line of force is not zero. The order of contact with the common tangent line of the line of force will be two or more if $\lambda = 3$. Thus we may state the following result.

THEOREM 7. *The ratio ρ of the curvatures at E_0 of the rest trajectory and the line of force is given by the formula*

$$\rho = (1 - \lambda)/(3 - \lambda) \cdot \tag{35}$$

For an ordinary field of force, the angular rate $\lambda = 0$ so that $\rho = 1/3$. There are generalized fields of force for which the ratio $\rho = 1/3$. We shall determine these in a later section.

9. The generalization of the theorem concerning $1/(2n+1)$ of the rates of departure. In this section, we shall generalize our Theorem 7 to the case where the rest trajectory has contact of order $n \geqq 2$ with the line of the lineal element E_0 at the point O of E_0. For this purpose, it is found advisable to take the element E_0 in the normal form, that is, its point O at the origin and its direction that of the x-axis. Therefore our rest trajectory may be written in the form

$$y = \frac{c_{n+1}}{(n+1)!}x^{n+1} + \frac{c_{n+2}}{(n+2)!}x^{n+2} + \cdots. \tag{36}$$

The line of force has at least first order contact with the line of E_0. Let m denote its order of contact with this line. Therefore our line of force may be written as

$$y = \frac{C_{m+1}}{(m+1)!}x^{m+1} + \frac{C_{m+2}}{(m+2)!}x^{m+2} + \cdots. \tag{37}$$

The equation (36) must satisfy the differential equation (5). In order to perform this substitution, we need to know the expansion of the left-hand side of (29) when (36) is placed into it. For that purpose, it is necessary to know the total derivative of order p of the left-hand side of (29).

To solve this problem, it is convenient to introduce the following notation. If χ is any function of (x, y, y'), then

$$\chi^{(1)} = \left(\frac{\partial}{\partial x} + y'\frac{\partial}{\partial y}\right)\chi, \cdots, \chi^{(p)} = \left(\frac{\partial}{\partial x} + y'\frac{\partial}{\partial y}\right)^p \chi. \tag{38}$$

With this notation, we see that the first p derivatives of the left-hand side of (29) are

$$\begin{aligned}
&\frac{d}{dx}(\psi - y'\phi) = \psi^{(1)} - y'\phi^{(1)} + y''(\psi_{y'} - y'\phi_{y'} - \phi),\\
&\frac{d^2}{dx^2}(\psi - y'\phi) = \psi^{(2)} - y'\phi^{(2)} + y''a_{2,2} + y'''(\psi_{y'} - y'\phi_{y'} - \phi),\\
&\frac{d^3}{dx^3}(\psi - y'\phi) = \psi^{(3)} - y'\phi^{(3)} + y''a_{3,3} + y'''a_{3,2} + y^{IV}(\psi_{y'} - y'\phi_{y'} - \phi),\\
&\qquad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\frac{d^p}{dx^p}(\psi - y'\phi) = \psi^{(p)} - y'\phi^{(p)} + y''a_{p,p} + y'''a_{p,p-1} + y^{IV}a_{p,p-2}\\
&\qquad\qquad + \cdots + y^{(p)}a_{p,2} + y^{(p+1)}(\psi_{y'} - y'\phi_{y'} - \phi),
\end{aligned} \tag{39}$$

where $a_{p,q} = f_p(x, y, y', y'', \cdots, y^{(q)})$ only. This last formula may easily be proved by induction.

If the line of force has contact of order m with the line of E_0, not necessarily in the normal position, then from the preceding equations, it follows that

$$\begin{aligned}
&\psi^{(1)} - y'\phi^{(1)} = \psi^{(2)} - y'\phi^{(2)} = \cdots = \psi^{(m-1)} - y'\phi^{(m-1)} = 0,\\
&\psi^{(m)} - y'\phi^{(m)} + Y^{(m+1)}(\psi_{y'} - y'\phi_{y'} - \phi) = 0.
\end{aligned} \tag{40}$$

THEOREM 8. *The order of contact m of the line of force with the line of E_0 is equal to or greater than one less the order of contact n of the rest trajectory. That is, $m \geqq n-1$.*

Let us assume that $m < n$. Upon substituting the rest trajectory (36) into the expression $\psi - y'\phi$, we find that the power series expansion of this is (since $m < n$)

$$\psi - y'\phi = -\frac{[(\psi_{y'})_0 - \phi_0]C_{m+1}}{m!}x^m + \cdots. \tag{41}$$

Placing this and also the series expansion (36) of the rest trajectory into the differential equation (5), we find that the power of the term of lowest degree is $m+n-2$ (since the condition $m<n$ demonstrates that the term of degree $2n-2$ is higher). Hence upon placing the coefficient of x^{m+n-2} equal to zero, we find

$$-\frac{[(\psi_{y'})_0-\phi_0]c_{n+1}C_{m+1}(n-1-m)}{m!(n-1)!}=0. \tag{42}$$

Since the angular rate $\lambda\neq 1$, it follows from this equation that if $m<n$, then $m=n-1$. Hence in all possibilities $m\geqq n-1$. Therefore the order of contact of the line of force with the common tangent line can never be less than the order of contact of the rest trajectory decreased by one.

Now let ρ be the ratio of instantaneous departures of the rest trajectory and the line of force from their common tangent line. For the case $m=n-1$, obviously $\rho=0$. Consider next the case where $m=n$. Of course $c_{n+1}\neq 0$ and $C_{n+1}\neq 0$ for $m=n$, but $c_{n+1}\neq 0$ and $C_{n+1}=0$ for $m>n$. The power series expansion of $\psi-y'\phi$ when the rest trajectory is placed therein, is

$$\psi-y'\phi=\frac{[(\psi_{y'})_0-\phi_0][c_{n+1}-C_{n+1}]}{n!}x^n+\cdots. \tag{43}$$

Placing this and the series expansion (36) of the rest trajectory into our differential equation (5), and putting the coefficient of the term of lowest degree $2n-2$ equal to zero, we discover

$$\frac{c_{n+1}}{n!(n-1)!}\left[C_{n+1}\{(\psi_{y'})_0-\phi_0\}-c_{n+1}\{(\psi_{y'})_0-(2n+1)\phi_0\}\right]=0. \tag{44}$$

The ratio ρ of departures of the rest trajectory and the line of force from their common tangent line is defined by the ratio $\rho=c_{n+1}/C_{n+1}$. Therefore we derive the following result

THEOREM 9. *The ratio ρ of the departures from the line of the lineal element E_0 of the rest trajectory and the line of force is given by the formula*

$$\rho=(1-\lambda)/(2n+1-\lambda)\cdot \tag{45}$$

By Theorem 8, we note that $\rho=0$ if and only if $m=n-1$. By this formula $\rho=\infty$, that is $m>n$ if and only if the angular rate $\lambda=2n+1$. For any other case the orders of contact are the same, $m=n$, and ρ is a finite nonzero number.

For an ordinary field of force, the angular rate $\lambda=0$ so that $\rho=1/(2n+1)$. In the next section, we shall discuss all generalized fields of force for which the ratio $\rho=1/(2n+1)$.

If $\lambda\neq 2n+1$, then the order of contact m with the common tangent line

of the line of force is n, or $n-1$, where n is the order of contact of the rest trajectory. However, if $\lambda = 2n+1$, then the line of force may have higher order contact with the common tangent line. *This is a new phenomenon.* It can never arise in ordinary positional fields of force. This new situation is that the line of force can have the same order of contact with the common tangent line as that of the rest trajectory, or else one less, or higher.

10. **Generalized fields of force for which the ratio of the departures $\rho = 1/(2n+1)$.** By Theorem 9, we see that if $\rho = 1/(2n+1)$, then the angular rate $\lambda = 0$ at the lineal element E_0.

THEOREM 10. *All generalized fields of force for which the ratio of the departures $\rho = 1/(2n+1)$ must be such that the eliminant with respect to y' of the two equations*

$$\psi - y'\phi, \qquad \psi_{y'} - y'\phi_{y'} = 0, \tag{46}$$

is identically zero.

This theorem means geometrically that the force vector F_0 of the special lineal element E_0 (whose direction coincides with that of F_0) is tangent to the terminal curve C of the point O at the end point of F_0.

11. **Generalized velocity systems.** For our generalized fields of force, we define systems of curves which are generalizations of velocity systems for ordinary positional fields of force. (For the positional case, see the Princeton Colloquium, pp. 42–44.)

Consider a particle of unit mass in any generalized positional field of force with rectangular components ϕ and ψ. The equations of motion are then given by (1). If the initial position and the initial velocity are given, the motion is determined. If only the initial position and the direction of the motion are given, the radius of curvature r will depend for its value on the initial speed. Hence, in addition to the usual formula for the speed v, there must be a formula expressing v^2 in terms of x, y, y', r. This is furnished by the first of equations (3). This equation may be written in the form

$$v^2 = (\psi - y'\phi)(1 + y'^2)/y''. \tag{47}$$

In the actual trajectory v varies from point to point. If now we replace v^2 in this result by some constant, say $1/c$, the resulting equation may be written in the form

$$y'' = c(\psi - y'\phi)(1 + y'^2). \tag{48}$$

The curves satisfying this differential equation, they are not in general trajectories, we define as *generalized velocity systems*. For any generalized field of force, a curve is a velocity curve corresponding to the speed v_0, provided a particle starting from any lineal element of the curve with that speed describes a trajectory osculating the curve. In a given field of force there are ∞^3

trajectories and ∞^3 velocity curves. If c is given, we have ∞^2 velocity curves. For unit velocity, we have the following result.

THEOREM 11. *Any system of ∞^2 curves may be regarded as the totality of velocity curves corresponding to unit velocity in $\infty^{f(3)}$ generalized fields of force.*

Upon eliminating the constant c, we obtain the complete system of ∞^3 velocity curves for a given generalized field of force. The differential equation of all velocity curves is

$$(\psi - y'\phi)y''' = [\psi_x + y'(\psi_y - \phi_x) - y'^2\phi_y]y'' + [\psi_{y'} - y'\phi_{y'} - 3\phi + 2(\phi + y'\psi)/(1 + y'^2)]y''^2. \tag{49}$$

Therefore a complete velocity system is of the type (G), thus possessing the Property I.

Observe that the expression $(\psi - y'\phi)(1+y'^2)$ of equation (48) can represent any function of (x, y, y'). This result shows that any complete velocity system is identical with curvature trajectories(4). For an arbitrary family of ∞^2 curves, a curve is a *curvature trajectory* if at any lineal element E of this curve, its curvature is c times the curvature of the curve of the family which passes through E. For a given family of ∞^2 curves, there are ∞^3 curvature trajectories. These possess an additional differential property in addition to the Property I.

THEOREM 12. *Any system of ∞^3 curvature trajectories may represent the velocity curves of $\infty^{1+f(3)}$ generalized fields of force.*

By equations (6), it is seen that the velocity systems are the curvature trajectories

$$y''' = (\alpha_x + y'\alpha_y)y'' + \alpha_{y'}y''^2. \tag{50}$$

By this, it is seen that for a given system of ∞^3 curvature trajectories, the function α is completely determined except for an additive constant. Since β is entirely arbitrary, we therefore discover the validity of our theorem.

12. **The generalized systems S_k defined by $P = kN$.** In a generalized field of force, we may define systems of curves S_k which are generalizations of the corresponding systems S_k for the ordinary positional case (see the Princeton Colloquium, pp. 91–94).

If an arbitrary curve is drawn in a generalized field of force, and the particle of unit mass is started along it from one of its points with a given speed, the constrained motion along the given curve is determined. The acceleration along the curve is given by T, the tangential component of the force vector. So the speed at any point is determined by

$$v^2 = 2\int T ds. \tag{51}$$

The pressure P (of course normal to the curve, since the curve is considered smooth) is given by the formula

$$P = v^2/r - N. \tag{52}$$

If we increase the initial speed, the effect is to increase v^2 by a constant c; and hence P changes by the addition of a term c/r.

If the given curve is a trajectory, the initial speed may be so chosen that the pressure vanishes throughout the motion; that is, trajectories may be defined as curves of no constraint. Of course if a different initial speed is used, P will be of the form c/r; but, as regards the curves, they are completely characterized by $P=0$.

The general problem that we wish to consider is to find curves so that the pressure P shall be proportional to the normal component N of the force vector. So $P=kN$. To a given value of k there correspond ∞^3 such curves. The system so obtained will be denoted by S_k.

We note that $k=0$ gives the system S_0 of trajectories and $k=\infty$ gives the system S_∞ of velocity curves.

To derive the differential equation of the system S_k, we must eliminate the speed v from the equations

$$v^2 = (k+1)rN, \qquad vv_s = T. \tag{53}$$

The result is

$$Nr_s = nT - rN_s, \tag{54}$$

where $n=2/(k+1)$.

Expanding this, we discover as the differential equation of the system S_k

$$\begin{aligned}(\psi - y'\phi)y''' = {} & [\psi_x + y'(\psi_y - \phi_x) - y'^2\phi_{y'}]y'' \\ & + [\psi_{y'} - y'\phi_{y'} - 3\phi - (n-2)(\phi + y'\psi)/(1+y'^2)]y''^2.\end{aligned} \tag{55}$$

We note that $n=2$ gives the generalized dynamical trajectories, and $n=0$ gives the velocity curves.

THEOREM 13. *Any system S_k is represented by a differential equation of the type* (G) *characterized by the geometric Property* I. *Conversely, any system of ∞^3 curves with the Property* I *can represent the system S_k with $k\neq\infty$ of $\infty^{I(2)}$ generalized fields of force.*

In terms of the functions α and β defined by equations (6), our system S_k is represented by the differential equation of the type (G)

$$y''' = (\alpha_x + y'\alpha_y)y'' + (\alpha_{y'} + (n/2)\beta)y''^2. \tag{56}$$

Thence if $n\neq 0$, the functions α and β may be determined up to an arbitrary function of two variables.

We note that this theorem is not valid for velocity curves $k=\infty$ or $n=0$. In addition to Property I, the type (G) must possess the additional differential conditions of second order

$$
\begin{aligned}
G_{yy'} &= 2H_y + H_{xy'} + y'H_{yy'}, \\
G_{xy'} + y'G_{yy'} - G_y &= H_{xx} + 2y'H_{xy} + y'^2H_{yy'}.
\end{aligned} \tag{57}
$$

If these conditions are satisfied, then our differential equation of the type (G) represents curvature trajectories. These may therefore represent the velocity curves of $\infty^{1+f(3)}$ generalized fields of force.

13. **Curves of constant pressure.** We now define curves of constant pressure in a generalized field of force analogous to those in an ordinary positional field of force. (See the Princeton Colloquium, pp. 97–98.) Consider the curves corresponding to $P=c$, where c denotes any constant. The curves obtained may be termed curves of constant pressure. Only along such a curve is a constrained motion of a particle possible such that the pressure against the curve remains constant.

For a given value of c, a system of ∞^3 curves is obtained, whose intrinsic equation, found by differentiating the relation

$$P = v^2/r - N = c, \tag{58}$$

is

$$(c + N)r_s = 2T - rN_s. \tag{59}$$

By this, we see that this system is given by a differential equation of the type (G) and therefore possesses the geometric Property I.

THEOREM 14. *For a given value of c, there are ∞^3 pressure curves $P=c$ whose differential equation is of the type* (G) *characterized by the geometric Property* I. *Conversely, any system of ∞^3 curves with the Property* I *may represent the pressure curves for a preassigned c of $\infty^{f(2)}$ generalized fields of force.*

This result may be deduced from the differential equation of the pressure curves $P=c$, which may be written by means of equations (6) in the form

$$
\begin{aligned}
[1 + c(1 + y'^2)^{3/2}]y''' &= [\alpha_x + y'\alpha_y]y'' \\
&\quad + [\beta + \alpha_{y'} + 3cy'(1 + y'^2)^{1/2}e^{-\alpha}]y''^2.
\end{aligned} \tag{60}
$$

Thence for a preassigned c, α is determined up to an arbitrary function of two variables and β is thereby completely known.

If all systems S_k are combined, we obtain ∞^4 curves. Similarly combining all pressure curves, we find ∞^4 curves. Each of these *two distinct* quadruply-infinite families of curves is given by a differential equation of the form

$$y^{IV} = A y'''^2 + B y''' + C, \tag{61}$$

where A, B, C are functions of the curvature element (x, y, y', y''). (See the Princeton Colloquium, pp. 95, 98, 113–115.)

The system (61) may be characterized geometrically as follows. If for each of the ∞^1 curves of this family which pass through a given curvature element (x, y, y', y''), we construct the osculating conic (five-point contact), the locus of the centers of these conics is a conic passing through the given point in the given direction.

Such quadruply-infinite systems of curves (61) were first encountered in connection with the systems of extremals in a calculus of variation problem[9]. These differential equations also arise in connection with the study of special types of fields of force which depend upon the position of the point and the time only, and in many other connections[10].

In conclusion, we note that all our results are valid under the total projective group consisting of collineations and correlations. The only contact transformations of the plane which preserve differential equations of the type (G) are the projective ones.

---

[9] Kasner, *Systems of extremals in the calculus of variations*, Bull. Amer. Math. Soc. vol. 13 (1907) p. 290.

[10] See Kasner, Revista de Matematicas vol. 3 (1942) pp. 7–12.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.
ILLINOIS INSTITUTE OF TECHNOLOGY,
CHICAGO, ILL.

A PROPERTY OF THE ZEROS OF LEGENDRE POLYNOMIALS

BY

A. ZYGMUND

§1

1. Let m and n be any two positive integers such that

$$m/n \geqq 1 + \delta, \tag{1.1}$$

where δ is positive. It has been proved that for any trigonometric polynomial

$$T(\theta) = (1/2)a_0 + \sum_{\nu=1}^{n} (a_\nu \cos \nu\theta + b_\nu \sin \nu\theta) \tag{1.2}$$

of order n,

$$\max_{0 \leq \theta \leq 2\pi} |T(\theta)| \leqq A'_\delta \max_{\mu=0,1,\cdots,2m} \left| T\left(\frac{2\pi\mu}{2m+1}\right) \right|, \tag{1.3}$$

where the coefficient A'_δ depends on δ only(1).

Similarly, for $r \geqq 1$

$$\left\{ \frac{1}{2\pi} \int_0^{2\pi} |T(\theta)|^r \, d\theta \right\}^{1/r} \leqq B'_\delta \frac{1}{2m+1} \left\{ \sum_{\mu=0}^{2m} \left| T\left(\frac{2\pi\mu}{2m+1}\right) \right|^r \right\}^{1/r} \tag{1.4}$$

with B'_δ depending on δ only.

It is the main purpose of this note to show that there exist analogous results for ordinary polynomials

$$Q(x) = c_0 + c_1 x + \cdots + c_n x^n, \tag{1.5}$$

if instead of the points $2\pi\mu/(2m+1)$ we consider the zeros of the Legendre polynomial $P_m(x)$, where m satisfies the inequality (1.1). These zeros will be denoted by (2)

Presented to the Society, December 27, 1942; received by the editors December 1, 1942.

(1) See S. Bernstein, *Sur une classe de formules d'interpolation*, Bull. Acad. Sci. URSS. (7) vol. 4 (1931) pp. 1151–1161. A proof of (1.3) (as well as that of inequality (1.4) below) was also given by J. Marcinkiewicz and A. Zygmund, *Mean values of trigonometric polynomials*, Fund. Math. vol. 28 (1937) pp. 131–166, especially p. 148. The authors overlooked the earlier paper of Bernstein.

(2) Of course, the inequalities (1.3) and (1.4) for trigonometric polynomials may be easily translated into analogous results for ordinary polynomials, if we take for the $x_\nu^{(m)}$ the zeros of Tchebyscheff's polynomials. The case of Legendre abscissas requires, however, an entirely different argument.

$$x_1^{(m)}, x_2^{(m)}, \cdots, x_m^{(m)}. \tag{1.6}$$

By $\gamma_1^{(m)}, \gamma_2^{(m)}, \cdots, \gamma_m^{(m)}$ we shall denote the Christoffel numbers, that is the numbers such that

$$\int_{-1}^{+1} R(x)dx = \sum_{\mu=1}^{m} R(x_\mu^{(m)})\gamma_\mu^{(m)}, \tag{1.7}$$

for any polynomial $R(x)$ of degree not exceeding $2m-1$. It is very well known that these numbers are positive(3).

THEOREM 1. *Suppose that m and n are positive integers, $n<m$, and that r is any real number not less than 1. Then for any polynomial $Q(x)$ of degree n,*

$$\max_{-1\leq x\leq +1} |Q(x)| \leqq K_{m,n} \max_{1\leq \mu\leq m} |Q(x_\mu^{(m)})|, \tag{1.8}$$

$$\left\{\int_{-1}^{+1} |Q(x)|^r dx\right\}^{1/r} \leqq K_{m,n}\left\{\sum_{\mu=1}^{m} |Q(x_\mu^{(m)})|^r \gamma_\mu^{(m)}\right\}^{1/r}, \tag{1.9}$$

where $K_{m,n}$ depends on m and n only. If m and n satisfy (1.1) *then*

$$K_{m,n} \leqq A_\delta \tag{1.10}$$

with A_δ depending only on δ.

2. The proof of the inequalities (1.8) and (1.9) has certain features in common with that of (1.3) and (1.4). In the latter case, the main idea of the proof was that to a partial sum of a trigonometric Fourier series we may add a nonoverlapping group of terms so that the resulting expression is a simple linear combination of the first arithmetic means of the Fourier series. In the proof of (1.8) and (1.9) we shall apply a similar device to Fourier-Legendre series, but instead of the first we shall consider the second arithmetic means of the series.

Let us consider an arbitrary series $u_0+u_1+\cdots+u_n+\cdots$. The partial sums and the second Cesàro means of the series we shall denote by s_n and σ_n'', respectively. It is easy to see that, given three non-negative integers $n<p<q$, we can always find constants A, B, C (depending on n, p, q only) such that

$$A\sigma_n'' + B\sigma_p'' + C\sigma_q'' - s_{n+2} \tag{2.1}$$

contains only terms u_ν with $\nu>n+2$.

For the proof of this remark, we observe that

---

(3) See, for example, G. Szegö, *Orthogonal polynomials*, Amer. Math. Soc. Colloquium Publications vol. 23 (1939) p. 46 ff.

$$\sigma_n'' = \sum_{\nu=0}^{n}\left(1-\frac{\nu}{n+1}\right)\left(1-\frac{\nu}{n+2}\right)u_\nu = \sum_{\nu=0}^{n+2}\left(1-\frac{\nu}{n+1}\right)\left(1-\frac{\nu}{n+2}\right)u_\nu,$$

and that similar formulas hold for σ_p'' and σ_q''. This gives for the determination of the numbers A, B, C the equations

$$A+B+C=1,$$

$$A\left(\frac{1}{n+1}+\frac{1}{n+2}\right)+B\left(\frac{1}{p+1}+\frac{1}{p+2}\right)+C\left(\frac{1}{q+1}+\frac{1}{q+2}\right)=0,$$

$$\frac{A}{(n+1)(n+2)}+\frac{B}{(p+1)(p+2)}+\frac{C}{(q+1)(q+2)}=0.$$

Hence

$$(2.2)\quad A=\frac{(n+1)(n+2)}{(p-n)(q-n)},\quad C=\frac{(q+1)(q+2)}{(q-p)(q-n)},\quad B=-\frac{(p+1)(p+2)}{(q-p)(p-n)}\text{(4)}.$$

The next remark we need is the following theorem of Fejér(5): *The second Cesàro means of the series*

$$(2.3)\qquad \sum_{\nu=0}^{\infty}(\nu+1/2)P_\nu(x)P_\nu(t)$$

are non-negative for $-1\leqq x\leqq+1,\ -1\leqq t\leqq+1$.

Let us denote these means by $K_n''(x,t)$ and let $\sigma_n''(x;f)$ denote the second Cesàro means of the Fourier-Legendre series

$$(2.4)\qquad \sum_{\nu=0}^{\infty}a_\nu P_\nu(x)$$

of a function $f(x)$ defined in the interval $-1\leqq x\leqq+1$. Then

$$\sigma_n''(x;f)=\int_{-1}^{+1}f(t)K_n'(x,t)dt.$$

Hence

$$(2.5)\qquad \max_{-1\leqq x\leqq+1}|\sigma_n'(x;f)|\leqq\max_{-1\leqq t\leqq+1}|f(t)|\cdot\int_{-1}^{+1}K_n''(x,t)dt$$
$$=\max_{-1\leqq x\leqq+1}|f(x)|.$$

(4) Thus the numbers $A, -B, C$ are positive.

(5) L. Fejér, *Über die Laplacesche Reihe*, Math. Ann. vol. 67 (1909) pp. 76–109.

Again, if $r \geqq 1$ and $1/r+1/r'=1$, then

$$|\sigma_n''(x;f)| \leqq \int_{-1}^{+1} |f(t)|\, K_n''^{1/r} K_n^{1/r'}\, dt$$

$$\leqq \left\{\int_{-1}^{+1} |f(t)|^r K_n''(x,t)dt\right\}^{1/r} \left\{\int_{-1}^{+1} K_n''(x,t)dt\right\}^{1/r'},$$

so that

$$|\sigma_n''(x;f)|^r \leqq \int_{-1}^{+1} |f(t)|^r K_n'(x,t)dt. \tag{2.6}$$

Integrating both sides of the inequality over the interval $-1 \leqq x \leqq +1$, and inverting the order of integration on the right, we get

$$\int_{-1}^{+1} |\sigma_n''(x;f)|^r dx \leqq \int_{-1}^{+1} |f(t)|^r dt. \tag{2.7}$$

3. We shall now prove the following theorem which is in a sense a converse of the inequality (1.9), and which gives slightly more than we actually need for the proof of Theorem 1.

LEMMA. *Let $r \geqq 1$ and let m and n be two positive integers such that $n \leqq 2m-1$. Then, for any polynomial $R(x)$ of degree n,*

$$\left\{\sum_{\nu=1}^{m} |R(x_\nu^{(m)})|^r \gamma_\nu^{(m)}\right\}^{1/r} \leqq L_{m,n}\left\{\int_{-1}^{+1} |R(x)|^r dx\right\}^{1/r}, \tag{3.1}$$

where the coefficient $L_{m,n}$ depends on m and n only. If

$$n(1+\delta) \leqq 2m-1, \qquad \delta > 0, \tag{3.2}$$

then

$$L_{m,n} \leqq L(1+1/\delta^2), \tag{3.3}$$

where L is an absolute constant.

Proof. Let μ be any integer such that $n-2<\mu<2m-1$, and let A^*, B^*, C^* denote the numbers defined by the formulas (2.2) with n, p, q replaced by $n-2$, μ, $2m-1$, respectively(6). Hence, treating $R(x)$ as a Fourier-Legendre series, we get

$$R(x) = A^*\sigma_{n-2}''(x;R) + B^*\sigma_\mu''(x;R) + C^*\sigma_{2m-1}''(x;R). \tag{3.4}$$

Now let λ be any positive integer not exceeding $2m-1$, and let us consider the inequality (2.6) with f replaced by R and n by λ. We get

(6) The argument is valid even if $n=0$ or $n=1$, provided we set $\sigma_{-1}''=\sigma_{-2}''=0$.

$$\sum_{\nu=1}^{m} |\sigma_\lambda''(x_\nu^{(m)}; R)|^r \gamma_\nu^{(m)} \leqq \int_{-1}^{+1} |R(t)|^r \left\{ \sum_{\nu=1}^{m} \gamma_\nu^{(m)} K_\lambda''(x_\nu^{(m)}, t) \right\} dt \tag{3.5}$$

$$= \int_{-1}^{+1} |R(t)|^r dt,$$

since

$$\sum_{\nu=1}^{m} \gamma_\nu^{(m)} K_\lambda''(x_\nu^{(m)}, t) = \int_{-1}^{+1} K_\lambda''(x, t) dt = 1.$$

Hence, on account of (3.4) and using Minkowski's inequality,

$$\left\{ \sum_{\nu=1}^{m} |R(x_\nu^{(m)})|^r \gamma_\nu^{(m)} \right\}^{1/r} \leqq A^* \left\{ \sum_{\nu=1}^{m} |\sigma_{n-2}''(x_\nu^{(m)}; R)|^r \gamma_\nu^{(m)} \right\}^{1/r}$$

$$+ |B^*| \left\{ \sum_{\nu=1}^{m} |\sigma_\mu''(x_\nu^{(m)}; R)|^r \gamma_\nu^{(m)} \right\}^{1/r}$$

$$+ C^* \left\{ \sum_{\nu=1}^{m} \sigma_{2m-1}''(x_\nu^{(m)}; R)^r \gamma_\nu^{(m)} \right\}^{1/r}$$

$$\leqq (A^* + |B^*| + C^*) \left\{ \int_{-1}^{+1} |R(t)|^r dt \right\}^{1/r}.$$

This proves (3.1) with

$$L_{m,n} = A^* + |B^*| + C^*. \tag{3.6}$$

In order to prove (3.3) we assume that we have equality in (3.2), for this only increases δ, and we set

$$\mu = [((n-2) + (2m-1))/2] = m - 1 + [(n-1)/2].$$

Then

$$2m - 1 - (n-2) > 2m - 1 - n = n\delta,$$

$$2m - 1 - \mu = m - [(n-1)/2] \geqq m - (1/2)(n-1)$$

$$> (1/2)(2m - 1 - n) = (1/2)n\delta,$$

$$\mu - (n-2) = m + [(n-1)/2] - (n-1) \geqq m - (1/2)(n-1) - 1$$

$$= (1/2)(2m - 1 - n) = (1/2)n\delta.$$

Hence the denominators in (2.2), with n, p, q replaced by $n-2, \mu, 2m-1$, respectively, all exceed $(1/4)n^2\delta^2 = (1/4)(2m-1)^2\delta^2(1+\delta)^{-2}$. The numerators are $O(m^2)$. From this and (3.6) we deduce (3.3).

4. We now pass to the proof of (1.9). We note that

$$\left\{ \int_{-1}^{+1} |Q(x)|^r dx \right\}^{1/r} = \max_f \int_{-1}^{+1} Q(x) f(x) dx, \tag{4.1}$$

for all f with

$$\left\{\int_{-1}^{+1}|f|^{r'}dx\right\}^{1/r'} \leqq 1. \tag{4.2}$$

For f we may take continuous functions only, or even only polynomials. We now observe that, if $s_n(x;f)$ denotes the nth partial sum of the Fourier-Legendre series of $f(x)$, then

$$\int_{-1}^{+1}Q(x)f(x)dx = \int_{-1}^{+1}Q(x)s_n(x;f)dx = \int_{-1}^{+1}Q(x)T(x)dx, \tag{4.3}$$

where $T(x)$ is any polynomial such that the nth partial sum of its Fourier-Legendre development coincides with $s_n(x;f)$. Let us set

$$T(x) = A'\sigma''_{n-2}(x;f) + B'\sigma''_p(x;f) + C'\sigma''_m(x;f), \tag{4.4}$$

where p is any integer satisfying $n-2<p<m$, and the numbers A', B', C' are defined by (2.2) with n, p, q replaced by $n-2$, p, m, respectively. Since $Q(x)\ T(x)$ is of degree $n+m\leqq 2m-1$,

$$\begin{aligned}\int_{-1}^{+1}Q(x)T(x)dx &= \sum_{\nu=1}^{m}Q(x_\nu^{(m)})T(x_\nu^{(m)})\gamma_\nu^{(m)} \\ &\leqq \left\{\sum_{\nu=1}^{m}|Q(x_\nu^{(m)})|^r\gamma_\nu^{(m)}\right\}^{1/r}\left\{\sum_{\nu=1}^{m}|T(x_\nu^{(m)})|^{r'}\gamma_\nu^{(m)}\right\}^{1/r'}.\end{aligned} \tag{4.5}$$

Applying Lemma 1 to the last sum and denoting by L' the least upper bound of the sequence $L_{m,m}$ (L' is finite, on account of (3.3)), we get

$$\begin{aligned}&\left\{\sum_{\nu=1}^{m}|T(x_\nu^{(m)})|^{r'}\gamma_\nu^{(m)}\right\}^{1/r'} \\ &\quad\leqq L_{m,m}\left\{\int_{-1}^{+1}|T(x)|^{r'}dx\right\}^{1/r'} \leqq L'\left\{\int_{-1}^{+1}|T(x)|^{r'}dx\right\}^{1/r'} \\ &\quad\leqq L'\left[A'\left\{\int_{-1}^{+1}|\sigma''_{n-2}|^{r'}dx\right\}^{1/r'} + |B'|\left\{\int_{-1}^{+1}|\sigma''_p|^{r'}dx\right\}^{1/r'}\right. \\ &\qquad\left. + C'\left\{\int_{-1}^{+1}|\sigma''_m|^{r'}dx\right\}^{1/r'}\right] \\ &\quad\leqq L'(A'+|B'|+C')\left\{\int_{-1}^{+1}|f|^{r'}dx\right\}^{1/r'} \leqq K_{m,n},\end{aligned}$$

with

$$K_{m,n} = L'(A' + |B'| + C'). \tag{4.6}$$

From this, (4.1), (4.3), and (4.5) we deduce (1.9). It may be added that, if $r=1$, the above argument does not require Lemma 1. What it uses is the trivial inequality.

$$\max_{\nu=1,\cdots,m} |Q(x_\nu^{(m)})| \leqq \max_{-1\leqq x\leqq +1} |Q(x)|$$

(which of course is contained in (3.1) for $r=+\infty$).

The proof of the inequality (1.8) is essentially the same as that of (1.9). Instead of (4.1) and (4.2) we use the relation

$$\max_{-1\leqq x\leqq +1} |Q(x)| = \max_{f} \int_{-1}^{+1} Q(x)f(x)dx, \quad \text{for} \quad \int_{-1}^{+1} |f|\,dx \leqq 1.$$

The rest of the argument is unchanged.

5. In order to prove (1.10), we take

$$p = [((n-2)+m)/2] = [(n+m)/2] - 1.$$

Then (the argument is analogous to (3.7))

$$\begin{aligned} m-(n-2) &> m-n \geqq n\delta, \\ m-p = m-[(n+m)/2]+1 &> m-(n+m)/2 \\ &= (1/2)(m-n) \geqq (1/2)n\delta, \\ p-(n-2) = [(m+n)/2]+1-n &> (m+n)/2-n \\ &= (1/2)(m-n) \geqq (1/2)n\delta, \end{aligned}$$

so that the denominators of A', B', C' are not less than $(1/4)n^2\delta^2=(1/4)m^2\cdot(1+\delta)^{-2}\delta^2$, assuming, as we may, that $m=n(1+\delta)$. Since the numerators are $O(m^2)$, we see that, assuming (1.1)

$$K_{m,n} \leqq K(1+\delta^{-2}), \tag{5.1}$$

where K is an absolute constant. This concludes the proof of Theorem 1.

§2

6. The argument used in §1 may be applied to some other problems. As an example, we shall prove the following:

THEOREM 2. *Let $r\geqq 1$ and let m and n be any two positive integers such that $m\geqq n$. Then every polynomial $R(x)$ of degree n may be represented in the form*

$$R(x) = R_1(x) - R_2(x), \tag{6.1}$$

where R_1 and R_2 are polynomials of degrees at most m, are non-negative for

$-1 \leqq x \leqq +1$, *and satisfy the inequalities*

$$\left\{\int_{-1}^{+1} R_i^r(x)dx\right\}^{1/r} \leqq K_{m,n}\left\{\int_{-1}^{+1} |R(x)|^r dx\right\}^{1/r} \tag{6.2}$$

with $K_{m,n}$ depending on m and n only. If m and n satisfy (1.1), *then*

$$K_{m,n} \leqq K(1+\delta^{-2}) \tag{6.3}$$

where K is an absolute constant.

Proof. Let us set

$$R(x) = R^+(x) - R^-(x),$$

where

$$R^+(x) = \max\{R(x), 0\}, \quad R^-(x) = \max\{-R(x), 0\},$$

so that

$$0 \leqq R^+(x) \leqq |R(x)|, \qquad 0 \leqq R^-(x) \leqq |R(x)|,$$

$$\int_{-1}^{+1}\{R^+(x)\}^r dx \leqq \int_{-1}^{+1}|R(x)|^r dx, \qquad \int_{-1}^{+1}\{R^-(x)\}^r dx \leqq \int_{-1}^{+1}|R(x)|^r dx.$$

Let p be any integer satisfying the inequalities

$$n-2 < p < m$$

and let A', B', C' be defined by the formulas (2.2) with n, p, q replaced by $n-2$, p, m, respectively. Let us consider the expressions

$$T^*(x) = A'\sigma''_{n-2}(x; R^+) + B'\sigma''_p(x; R^+) + C'\sigma''_m(x; R^+), \tag{6.4}$$

$$T_*(x) = A'\sigma''_{n-2}(x; R^-) + B'\sigma''_p(x; R^-) + C'\sigma''_m(x; R^-) \tag{6.5}$$

whose difference is

$$A'\sigma''_{n-2}(x; R) + B'\sigma''_p(x; R) + C'\sigma''_m(x; R) = R(x),$$

and let us set

$$R_1(x) = A'\sigma''_{n-2}(x; R^+) - B'\sigma''_p(x; R^-) + C'\sigma''_m(x; R^+),$$

$$R_2(x) = A'\sigma''_{n-2}(x; R^-) - B'\sigma''_p(x; R^+) + C'\sigma''_m(x; R^-).$$

The polynomials R_1 and R_2 are non-negative over the interval $-1 \leqq x \leqq +1$ and their degrees do not exceed m. Clearly (6.1) is satisfied, since

$$R_1(x) - R_2(x) = T^*(x) - T_*(x) = R(x).$$

Moreover, using Minkowski's inequality, and taking into account (2.7), we get

$$\left\{\int_{-1}^{+1} R_1^r dx\right\}^{1/r} \leqq A'\left\{\int_{-1}^{+1} \sigma_{n-2}'^{r}(x; R^+)dx\right\}^{1/r} + |B'|\left\{\int_{-1}^{+1} \sigma_p'^{r}(x; R^-)dx\right\}^{1/r} + C'\left\{\int_{-1}^{+1} \sigma_m'^{r}(x; R^+)dx\right\}^{1/r} \leqq K_{m,n}\left\{\int_{-1}^{+1} |R(x)|^r dx\right\}^{1/r},$$

with

$$K_{m,n} = A' + |B'| + C'.$$

Similarly we prove (6.2) for $i=2$.

A comparison of this argument with that of section 4 shows that the constant $K_{m,n}$ of Theorem 2 is the same as the $K_{m,n}$ of Theorem 1. This immediately gives the inequality (6.2) (cf. (5.1)).

7. It is clear that an argument similar to that of the preceding section (and even slightly simpler, because we may use the first arithmetic means instead of the second) gives an analogue of Theorem 2 for trigonometric polynomials

$$U(\theta) = (1/2)a_0 + \sum_{\nu=1}^{n} (a_\nu \cos \nu\theta + b_\nu \sin \nu\theta). \tag{7.1}$$

THEOREM 3. *Let $r \geqq 1$ and let $n \leqq m$ be two positive integers. Every trigonometric polynomial* (7.1) *of order n may be represented in the form*

$$U(\theta) = U_1(\theta) - U_2(\theta), \tag{7.2}$$

where U_1 and U_2 are non-negative trigonometric polynomials of order at most m, satisfying the inequalities

$$\left\{\int_0^{2\pi} U_i^r d\theta\right\}^{1/r} \leqq K'_{m,n}\left\{\int_0^{2\pi} |U|^r d\theta\right\}^{1/r}, \tag{7.3}$$

with $K'_{m,n}$ depending on m and n only. If m and n satisfy (1.1), *then*

$$K'_{m,n} \leqq K'(1 + \delta^{-1}) \tag{7.4}$$

where K' is an absolute constant.

The proof may be condensed at some points. Given a function $f(\theta)$ of period 2π we shall denote by $s_n(\theta; f)$ and $\sigma_n'(\theta; f)$ the partial sums and the first arithmetic means of the Fourier series of f. If $p<q$ are two non-negative integers, then the linear combination

$$-\frac{n}{m-n+1}\sigma_{n-1}(\theta;f)+\frac{m+1}{m-n+1}\sigma_m(\theta;f),$$

written as a trigonometric polynomial, has its nth partial sum equal to $s_n(\theta;f)$. Similarly as in section 6 we introduce the functions $U^+(\theta)$ and $U^-(\theta)$ and the polynomials

$$U_1(\theta)=\frac{n}{m-n+1}\sigma_{n-1}(\theta;U^-)+\frac{m+1}{m-n+1}\sigma_m(\theta;U^+),$$

$$U_2(\theta)=\frac{n}{m-n+1}\sigma_{n-1}(\theta;U^+)+\frac{m+1}{m-n+1}\sigma_m(\theta;U^-).$$

These polynomials are non-negative, of degree at most m, and satisfy (7.2). Taking into account that, for any function $f(\theta)$,

$$\left\{\int_0^{2\pi}|\sigma_\mu(\theta;f)|^r d\theta\right\}^{1/r}\leqq\left\{\int_0^{2\pi}|f(\theta)|^r d\theta\right\}^{1/r},\qquad 1\leqq r\leqq+\infty,$$

we get (7.3) where

$$K'_{m,n}=\frac{m+n-1}{m-n+1}<\frac{2m+1}{m-n+1}.$$

This yields (7.4), with $K'=2$, provided that m and n satisfy (1.1)

§3

8. In the proof of Theorem 1, instead of the second arithmetic means $\sigma_n''(x;f)$ of the Fourier-Legendre series of the function f, we could have used the first arithmetic means $\sigma_n'(x;f)$ of that series. The argument would have been similar, for although the first arithmetic means $K_n'(x,t)$ of the series (2.3) are not positive, they have the property that the integral

$$\int_{-1}^{+1}|K_n'(x,t)|\,dt \tag{8.1}$$

is bounded in n and x(7), and that is the only property of the kernel that was required in the proof. The advantage of using the second arithmetic means is that the positiveness of the kernel $K_n''(x,t)$ is more elementary than the boundedness of the integral (8.1). Moreover, for the proof of Theorem 2 the fact that we used a positive kernel was essential. On the other hand, using the kernel $K_n'(x,t)$ we might obtain a better estimate for $K_{m,n}$ than the inequality (5.1). Namely, we might obtain that $K_{m,n}=O(\delta^{-1})$ for $\delta\to+0$. That this estimate is still too crude is shown by the following:

(7) See T. H. Gronwall, *Über die Laplacesche Reihe*, Math. Ann. vol. 74 (1913) pp. 213–270.

THEOREM 4. *Suppose that* $0<\delta\leqq 1$, *and that* A_δ *is the least number satisfying* (1.10). *Then*

$$A_\delta \leqq M\delta^{-1/2}, \tag{8.2}$$

where M *denotes an absolute constant.*

It may be observed that the order $\delta^{-1/2}$ in (8.2) cannot be improved. For if, for a given n, we take the least possible value of m, namely $m=n+1$, then $\delta=1/n$. On the other hand, the polynomial $P(x)$ is then determined uniquely by its values at the points (1.6). It is, however, well known that a polynomial $P(x)$ of degree n, which is absolutely less than 1 at the points (1.6) where $m=n+1$, may be of the order $n^{1/2}$ at some points of the interval $-1\leqq x\leqq +1$(8). Hence $\delta^{-1/2}$ is the right order of the coefficient A_δ.

In the proof of Theorem 3 it is slightly more preferable to use the kernel $K_n'(x, t)$ instead of $K_n''(x, t)$. However we shall not appeal to the fact that the integral (8.1) is bounded (on the contrary, our argument would easily prove that fact).

We shall confine our attention to the inequality (1.10). A practically identical argument gives the inequality (1.9), which also may be directly deduced from (1.10) by making r tend to $+\infty$.

Let us revert to the proof of Theorem 1 (see section 4). We start with the inequalities (4.1) and (4.2), where f denotes a polynomial. In the formula (4.3) we replace $T(x)$ by the polynomial

$$\tau(x) = -\frac{n}{m-n+1}\sigma_{n-1}'(x; f) + \frac{m+1}{m-n+1}\sigma_m'(x; f) \tag{8.3}$$

which is of degree at most m, and which has the property that

$$s_n(x; \tau) = s_n(x; f).$$

Instead of (4.5) we get

$$\int_{-1}^{+1} Q(x)\tau(x)dx \leqq \left\{\sum_{\nu=1}^{m} |Q(x_\nu^{(m)})|^r \gamma_\nu^{(m)}\right\}^{1/r} \left\{\sum_{\nu=1}^{m} |\tau(x_\nu^{(m)})|^{r'} \gamma_\nu^{(m)}\right\}^{1/r'}, \tag{8.4}$$

where, with the notation of section 4,

$$\left\{\sum_{\nu=1}^{m} |\tau(x_\nu^{(m)})|^{r'} \gamma_\nu^{(m)}\right\}^{1/r'} \leqq L'\left\{\int_{-1}^{+1} |\tau(x)|^{r'} dx\right\}^{1/r'}. \tag{8.5}$$

Theorem 4 will have been established, therefore, when we have shown that

$$\left\{\int_{-1}^{+1} |\tau(x)|^{r'} dx\right\}^{1/r'} \leqq M'\delta^{-1/2}, \tag{8.6}$$

where M' denotes an absolute constant.

(8) See, for example, G. Szegö, loc. cit. pp. 328–332.

9. We observe that

$$\tau(x) = \int_{-1}^{+1} f(t)H(x, t)dt, \tag{9.1}$$

where

$$H(x, t) = -\frac{n}{m-n+1} K'_{n-1}(x, t) + \frac{m+1}{m-n+1} K'_m(x, t). \tag{9.2}$$

We shall find for the integral

$$\int_{-1}^{+1} | H(x, t) | dt \tag{9.3}$$

a majorant independent of x. For this purpose let us consider the first arithmetic means of the series

$$\sum_{\nu=0}^{\infty} (\nu + 1/2)P_\nu(\cos \gamma). \tag{9.4}$$

These means will be denoted by $\chi_n(\gamma)$. Using the familiar formula

$$P_\nu(\cos \theta)P_\nu(\cos \theta') = \int_0^{2\pi} P_\nu(\cos \gamma)d\phi, \tag{9.5}$$

where $\cos \gamma = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos \phi$, and setting

$$x = \cos \theta, \qquad t = \cos \theta',$$

$$h(\gamma) = -\frac{n}{m-n+1} \chi_{n-1}(\gamma) + \frac{m+1}{m-n+1} \chi_m(\gamma),$$

we find that

$$\int_{-1}^{+1} | H(x, t) | dt = \int_0^{\pi} \left| \int_0^{2\pi} h(\gamma)d\phi \right| \sin \theta' d\theta' \leqq \iint_S | h(\gamma) | d\omega \tag{9.6}$$

where S denotes the unit sphere and $d\omega$ is the element of area on S. It is clear that the last integral, which we shall denote by I, is independent of θ, so that we may take for example $\theta = 0$. Thus $\theta' = \gamma$ and

$$I = 2\pi \int_0^{2\pi} | h(\gamma) | \sin \gamma d\gamma. \tag{9.7}$$

Let us now revert to the formula (9.1). Hölder's inequality gives

$$|\tau(x)| \leqq \left\{\int_{-1}^{+1} |f(t)|^{r'} |H(x,t)|\,dt\right\}^{1/r'} \left\{\int_{-1}^{+1} |H(x,t)|\,dt\right\}^{1/r}$$

$$\leqq I^{(r'-1)/r'} \left\{\int_{-1}^{+1} |f(t)|^{r'} |H(x,t)|\,dt\right\}^{1/r'},$$

so that

$$\int_{-1}^{+1} |\tau(x)|^{r'} dx \leqq I^{r'-1} \int_{-1}^{+1} dx \int_{-1}^{+1} |f(t)|^{r'} |H(x,t)|\,dt$$

$$= I^{r'-1} \int_{-1}^{+1} |f(t)|^{r'} dt \int_{-1}^{+1} |H(x,t)|\,dt \leqq I^{r'},$$

on account of (9.6). This fact and (8.6) show that Theorem 4 will be a consequence of the inequality

$$I \leqq M' \delta^{-1/2} \tag{9.8}$$

which we are now going to prove.

10. We shall use the fact, first pointed out by Fejér(9), that the series (9.4) is a Cauchy product of the two series

$$\sum_{\nu=0}^{\infty} P_\nu(\cos\gamma), \qquad 1/2 + \sum_{\nu=1}^{\infty} \cos\nu\gamma. \tag{10.1}$$

Taking into account the familiar formula of Mehler (10)

$$P_\mu(\cos\gamma) = \frac{2}{\pi} \int_\gamma^\pi \frac{\sin(\mu + 1/2)t}{[2(\cos\gamma - \cos t)]^{1/2}}\,dt,$$

we see that the νth partial sum of the first series (10.1) is equal to

$$\frac{2}{\pi} \int_\gamma^\pi \frac{1 - \cos(\nu+1)t}{2\sin(1/2)t\,[2(\cos\gamma - \cos t)]^{1/2}}\,dt = V_\nu,$$

say. Since the νth partial sum of the second series (10.1) is the Dirichlet kernel $D_\nu(\gamma) = \sin(\nu+1/2)\gamma/2 \sin(1/2)\gamma$, the first arithmetic means $\chi_N(\gamma)$ of the series (9.4) are given by the formula

$$(N+1)\chi_N(\gamma) = \sum_{\nu=0}^{N} V_\nu(\gamma) D_{N-\nu}(\gamma)$$

$$= \frac{2}{\pi} \cdot \frac{1}{2\sin\gamma/2} \int_\gamma^\pi \frac{\sum_{\nu=0}^{N} \sin(N-\nu+1/2)\gamma\,[1-\cos(\nu+1)t]}{2\sin(t/2)\,[2(\cos\gamma - \cos t)]^{1/2}}\,dt.$$

(9) Fejér, loc. cit.

(10) See, for example, G. Szegő, loc. cit. p. 87.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

Let us now represent here the products of sines by cosines as differences of sines, and let us apply the formulas for the sums of sines of an arithmetic progression. The last formula may then be written

$$(10.2)\qquad (N+1)\chi_N(\gamma) = \frac{2}{\pi}\cdot\frac{1}{2\sin\gamma/2}\int_\gamma^\pi \frac{\alpha_N^{(1)}(\gamma)+\alpha_N^{(2)}(\gamma,t)+\alpha_N^{(3)}(\gamma,t)}{2\sin(t/2)\,[2(\cos\gamma-\cos t)]^{1/2}}\,dt,$$

where

$$\alpha_N^{(1)}(\gamma) = \frac{1-\cos(N+1)\gamma}{2\sin\gamma/2},$$

$$\alpha_N^{(2)}(\gamma,t) = -\frac{1}{2}\cdot\frac{\cos(N+1)t-\cos(N+1)\gamma}{2\sin(t-\gamma)/2},$$

$$\alpha_N^{(3)}(\gamma,t) = \frac{1}{2}\cdot\frac{\cos(N+1)t-\cos(N+1)\gamma}{2\sin(t+\gamma)/2}.$$

Let us now consider the formula (9.5) and let us set

$$m-(n-1)=\lambda,$$

so that

$$(10.3)\qquad h(\gamma) = \lambda^{-1}[(m+1)\chi_m-(m-\lambda+1)\chi_{m-\lambda}].$$

Hence, if in (10.2) we substitute for N the values m and $m-\lambda$, and subtract the results, we get

$$h(\gamma) = h^{(1)}(\gamma)+h^{(2)}(\gamma)+h^{(3)}(\gamma),$$

where

$$(10.4)\qquad h^{(i)}(\gamma) = \frac{1}{\lambda}\cdot\frac{2}{\pi}\cdot\frac{1}{2\sin\gamma/2}\int_\gamma^\pi \frac{\alpha_m^{(i)}-\alpha_{m-\lambda}^{(i)}}{2\sin(t/2)\,[2(\cos\gamma-\cos t)]^{1/2}}\,dt \qquad (i=1,2,3).$$

We shall show that

$$(10.5)\qquad \int_0^\pi |h^{(i)}(\gamma)|\sin\gamma\,d\gamma \leqq C(m/\lambda)^{1/2},$$

where C is an absolute constant. (In the rest of this paper we shall denote by C absolute constants, not always the same at every occurrence.) Since the integral I (cf. (9.7)) does not exceed the sum of the integrals (10.5) multiplied by 2π, and since

$$\lambda > m-n \geqq n\delta \geqq m\delta/2,$$

so that $m/\lambda<2/\delta$, the inequalities (10.5) will imply (9.8).

11. Let us take for example $i=2$. We note that

$$|h^{(2)}(\gamma)|\sin\gamma \leqq 2(\pi\lambda)^{-1}\cos(\gamma/2)$$
$$\cdot\int_\gamma^\pi (2\sin t/2)^{-1}[2\sin(t-\gamma)/2\cdot 2\sin(t+\gamma)/2]^{-1/2}\,|\alpha_m^{(2)}-\alpha_{m-\lambda}^{(2)}|\,dt,$$

and we break up the integral on the right into three integrals extended respectively over the intervals $(\gamma, \gamma+1/m)$, $(\gamma+1/m, \gamma+1/\lambda)$, $(\gamma+1/\lambda, \pi)$.(11) Correspondingly

$$|h^{(2)}(\gamma)|\sin\gamma \leqq R_1(\gamma)+R_2(\gamma)+R_3(\gamma), \tag{11.1}$$

where

$$R_1=\lambda^{-1}\cos(\gamma/2)\int_\gamma^{\gamma+1/m},\qquad R_2=\lambda^{-1}\cos(\gamma/2)\int_{\gamma+1/m}^{\gamma+1/\lambda},$$
$$R_3=\lambda^{-1}\cos(\gamma/2)\int_{\gamma+1/\lambda}^{\pi}.$$

Observing that

$$\sin u \geqq (2/\pi)u \qquad (0\leqq u\leqq \pi/2),$$
$$\sin(\gamma+t)/2 \geqq \cos(\gamma/2)\sin(t/2) \qquad (0\leqq\gamma\leqq t\leqq\pi),$$

we find that

$$R_3(\gamma)\leqq C\lambda^{-1}\int_{\gamma+1/\lambda}^{\pi} t^{-3/2}(t-\gamma)^{-3/2}dt \leqq C\lambda^{1/2}\int_{1/\lambda}^{\infty} t^{-3/2}dt = C\lambda.$$

Again,

$$R_3(\gamma)\leqq C\lambda^{-1}\int_{\gamma+1/\lambda}^{\infty} t^{-3/2}(t-\gamma)^{-3/2}dt \leqq C\lambda^{-1}\gamma^{-3/2}\int_{\gamma+1/\lambda}^{\infty}(t-\gamma)^{-3/2}dt$$
$$= C\lambda^{-1/2}\gamma^{-3/2}.$$

Thus we got the following two inequalities for R_3:

$$R_3(\gamma)\leqq C\lambda,\qquad R_3(\gamma)\leqq C\lambda^{-1/2}\gamma^{-3/2}. \tag{11.2}$$

Taking into account the structure of the numerator of $\alpha^{(2)}$, we find that

$$R_2(\gamma)\leqq C\lambda^{-1}\int_{\gamma+1/m}^{\gamma+1/\lambda}\{|\sin\lambda t/2|+|\sin\lambda\gamma/2|\}t^{-3/2}(t-\gamma)^{-3/2}dt. \tag{11.3}$$

If we replace each sine on the right by 1, we get

(11) If any of these intervals partly projects outside $(0, \pi)$ we integrate over only the part contained in $(0, \pi)$.

$$R_2(\gamma) \leqq C\lambda^{-1}\int_{\gamma+1/m}^{\gamma+1/\lambda} t^{-3/2}(t-\gamma)^{-3/2}dt \leqq C\lambda^{-1}\gamma^{-3/2}\int_{\gamma+1/m}^{\infty}(t-\gamma)^{-3/2}dt$$
$$\leqq C\lambda^{-1}m^{1/2}\gamma^{-3/2}.$$

If we replace the sines by the corresponding angles and observe that $\gamma \leqq t$, we get

$$R_2(\gamma) \leqq C\int_{\gamma+1/m}^{\gamma+1/\lambda} t^{-1/2}(t-\gamma)^{-3/2}dt \leqq Cm^{1/2}\gamma^{-1/2}. \tag{11.4}$$

Thus

$$R_2(\gamma) \leqq Cm^{1/2}\gamma^{-1/2}, \qquad R_2(\gamma) \leqq C\lambda^{-1}m^{1/2}\gamma^{-3/2}. \tag{11.5}$$

Finally, we observe that

$$R_1(\gamma) \leqq C\lambda^{-1}\int_{\gamma}^{\gamma+1/m} t^{-3/2}(t-\gamma)^{-3/2}\left|\alpha_m^{(2)} - \alpha_{m-\lambda}^{(2)}\right| dt,$$

and that the numerator of $\alpha_m^{(2)}$ is

$$\sin (1/2)(m+1)(t-\gamma)\sin (1/2)(m+1)(t+\gamma),$$

so that the numerator of $\alpha_m^{(2)} - \alpha_{m-\lambda}^{(2)}$ is

$$\begin{aligned}
&[\sin (1/2)(m+1)(t-\gamma) - \sin (1/2)(m-\lambda+1)(t-\gamma)]\sin (1/2)(m+1)(t+\gamma)\\
&\quad + [\sin (1/2)(m+1)(t+\gamma) - \sin (1/2)(m-\lambda+1)(t+\gamma)]\\
&\quad \cdot \sin (1/2)(m-\lambda+1)(t-\gamma)\\
&= 2\sin (1/4)\lambda(t-\gamma)\cos (1/2)(m+1-\lambda/2)(t-\gamma)\\
&\quad \cdot \sin (1/2)(m+1)(t+\gamma)\\
&\quad + 2\sin (1/4)\lambda(t+\gamma)\cos (1/2)(m+1-\lambda/2)(t+\gamma)\\
&\quad \cdot \sin (1/2)(m-\lambda+1)(t-\gamma).
\end{aligned} \tag{11.6}$$

If we replace each sine on the right by the corresponding angle, the cosines by 1, and if we note that $\gamma + t \leqq 2t$, we find that

$$R_1(\gamma) \leqq Cm\int_{\gamma}^{\gamma+1/m} t^{-1/2}(t-\gamma)^{-1/2}dt \leqq C\gamma^{-1/2}m\int_{\gamma}^{\gamma+1/m}(t-\gamma)^{-1/2}dt$$
$$\leqq Cm^{1/2}\gamma^{-1/2}.$$

However, we may also note that the right-hand side of (11.6) does not exceed

$$C[\lambda(t-\gamma) + m(t-\gamma)] \leqq Cm(t-\gamma)$$

so that

$$R_1(\gamma) \leqq Cm\lambda^{-1} \int_{\gamma}^{\gamma+1/m} t^{-3/2}(t-\gamma)^{-1/2}dt \leqq Cm^{1/2}\lambda^{-1}\gamma^{-3/2}.$$

Thus

$$R_1(\gamma) \leqq Cm^{1/2}\gamma^{-1/2}, \qquad R_1(\gamma) \leqq Cm^{1/2}\lambda^{-1}\gamma^{-3/2}. \tag{11.7}$$

The inequalities we have proved show that

$$\int_0^{\pi} R_k(\gamma)d\gamma \leqq C(m/\lambda)^{1/2} \qquad (k = 1, 2, 3). \tag{11.8}$$

For $k=3$, this follows immediately if we break up the integral (11.8) into two, extended over the intervals $0 \leqq \gamma \leqq 1/\lambda$ and $1/\lambda \leqq \gamma \leqq \pi$, and apply respectively the inequalities (11.2):

$$\int_0^{\pi} R_3 d\gamma \leqq \int_0^{1/\lambda} C\lambda d\gamma + \int_{1/\lambda}^{\infty} C\lambda^{-1/2}\gamma^{-3/2}d\gamma \leqq C \leqq C(m/\lambda)^{1/2}.$$

Similarly, using the inequalities (11.4) and (11.7), we settle the cases $k=2$, 3 of (11.8). Taking into account the definition of the functions R_k we deduce from (11.8) the validity of (10.5) for $i=2$.

The case $i=3$ is similar to that of $i=2$, the only difference being that the factor $2 \sin (t-\gamma)/2$ in the denominator of $\alpha^{(2)}$ is replaced by $2 \sin (t+\gamma)/2$, which is larger, on account of the obvious inequality

$$(t-\gamma)/2 \leqq (t+\gamma)/2 \leqq \pi - (t-\gamma)/2.$$

Thus the integral (10.5) when $i=3$ is less than the integral with $i=2$.

The case $i=1$ is also easier than that of $i=2$ but is slightly different. We observe that

$$\begin{aligned} &\cos(\gamma/2)\int_{\gamma}^{\pi}(2\sin t/2)^{-1}[2(\cos\gamma - \cos t)]^{-1/2}dt \\ &\leqq C\left(\int_{\gamma}^{\gamma+1/\lambda} + \int_{\gamma+1/\lambda}^{\infty}\right) t^{-3/2}(t-\gamma)^{-1/2}dt \\ &\leqq C\left[\gamma^{-3/2}\int_{\gamma}^{\gamma+1/\lambda}(t-\gamma)^{-1/2}dt + \lambda^{1/2}\int_{\gamma+1/\lambda}^{\infty} t^{-3/2}dt\right] \\ &= C[\lambda^{-1/2}\gamma^{-3/2} + \lambda^{1/2}\gamma^{-1/2}]. \end{aligned} \tag{11.9}$$

Of the two terms in the last square brackets, the first prevails for $\gamma \leqq 1/\lambda$, the second for $\gamma \geqq 1/\lambda$. Hence, if we denote the left-hand side of (11.9) by $E(\gamma)$, we have the following inequalities

$$E(\gamma) \leqq C\lambda^{-1/2}\gamma^{-3/2}, \text{ if } \lambda\gamma \leqq 1; \quad E(\gamma) \leqq C\lambda^{1/2}\gamma^{-1/2}, \text{ if } \lambda\gamma \geqq 1. \tag{11.10}$$

Obviously

$$(11.11)\qquad \begin{aligned} |h^{(1)}(\gamma)| \sin\gamma &\leqq C\lambda^{-1}E(\gamma)\,|\cos(m+1)\gamma-\cos(m-\lambda+1)\gamma|\,/2\sin\gamma/2 \\ &\leqq C\lambda^{-1}E(\gamma)\,|\sin(\lambda\gamma/2)\sin(m+1-\lambda/2)\gamma|\,/2\sin\gamma/2. \end{aligned}$$

The right-hand side here may only be increased if the product of sines in the denominator is replaced by one of the following numbers: $\lambda(m+1-\lambda/2)\gamma^2/2$, $\lambda\gamma/2$, 1. The first will be the most advantageous in the interval $0\leqq\gamma\leqq 1/m$, the second for $1/m\leqq\gamma\leqq 1/\lambda$, the third for $\gamma\geqq 1/\lambda$. Taking into account (11.10), we deduce from (11.11) that $|h^{(1)}(\gamma)|\sin\gamma$ is majorized by the following expressions

$$Cm\lambda^{-1/2}\gamma^{-1/2},\quad \text{if}\quad 0\leqq\gamma\leqq 1/m;\qquad C\lambda^{-1/2}\gamma^{-3/2},\quad \text{if}\quad \gamma\geqq 1/m,$$

and these inequalities give (10.5) for $i=1$. This completes the proof of Theorem 4.

12. It may be of some interest to point out that the constants A_i' and B_i', of the inequalities (1.3) and (1.4) are $O(\log 1/\delta)$ as $\delta\to+0$. For the inequality (1.3) this is proved in Bernstein's paper cited above. The inequality for B_i' follows almost immediately from the argument used in the paper by Marcinkiewicz and Zygmund(1), provided we use the fact that, if

$$K_n(t) = (n+1)^{-1}\left\{\sin^2(n+1)t/2\right\}/2\sin^2 t/2$$

denotes Fejér's kernel, and $0<\lambda<m$, then

$$\int_0^\pi \lambda^{-1}\left|(m+1)K_{m-1}(t)-(m-\lambda+1)K_{m-\lambda}(t)\right|dt = O(1+\log m/\lambda).$$

For the proof of this inequality it is sufficient to split up this integral into three, extended respectively over the intervals $(0, 1/m)$, $(1/m, 1/\lambda)$, $(1/\lambda, \pi)$, and to observe that in the first integral the integrand is $O(m)$, in the second it is $O(1/t)$, and in the third $O(1/\lambda t^2)$.

MT. HOLYOKE COLLEGE,
SOUTH HADLEY, MASS.

FLAT SPACE CONGRUENCES OF ORDER ONE IN $[n]$

BY

EDWIN J. PURCELL

By an $[n-k]$-congruence of order one in $[n]$, $k<n$, is meant an algebraic ∞^k-system of $[n-k]$'s in n-dimensional projective space such that one and only one $[n-k]$ of the system passes through an arbitrary point of $[n]$. A point on more than one $[n-k]$ of the congruence is called a *fundamental point*. A fundamental point fails to determine any $[n-k]$ of the congruence.

Although some special cases of flat space congruences have been studied, there is no general theory in the literature(1). Line congruences of order one in $[3]$ were investigated by E. Kummer(2) and this was completed by R. Sturm(3). G. Marletta(4,5) classified line congruences of order one in $[4]$ but his methods were not suitable for extension to $[n]$. Some line congruences in $[5]$ were discussed by M. Sgroi(6). G. Aprile(7) collected the known examples of line congruences of order one in $[n]$; these are all special cases of the more general congruences defined in the present paper and it will be shown in §8 that his construction for a line congruence in $[n]$ with irreducible locus of fundamental points is a restricted case of one of ours(8).

Probably not all possible flat space congruences of order one are included in what follows. The line congruence in $[4]$ consisting of the trisecants of the F^4, projection of the Veronese surface in $[5]$, and the two line congruences in $[5]$ consisting of the bisecants of a ruled F^4 or of the bisecants of the del Pezzo surface, do not appear to belong to our class of congruences.

1. **Classification.** On each $[n-k]$ of our $[n-k]$-congruences of order one

Presented to the Society, November 28, 1942; received by the editors April 6, 1942, and, in revised form, January 6, 1943.

(1) Encyklopädie der Mathematischen Wissenschaften vol. III C 7, p. 964.

(2) E. Kummer, *Über die algebraischen Strahlensysteme, in besondere die der ersten und zweiten Ordnung*, Berlin, Abhandlungen der Preussischen Akademie der Wissenschaften, 1866.

(3) R. Sturm, *Die Gebilde ersten und zweiten Grades der Liniengeometrie in synthetischer Behandlung*, Leipzig, vol. 2, 1893.

(4) G. Marletta, *Sui complessi di rette del primo ordine dello spazio a quattro dimensioni*, Rend. Circ. Mat. Palermo vol. 28 (1909).

(5) Marletta, *Sopra i complessi di rette d'ordine uno dell S_4*, Atti Accademia Gioenia Catania, (5) vol. 3 (1909).

(6) M. Sgroi, *Sui 4-complessi di rette d'ordine uno nell'S_5*, Atti Accademia Gioenia Catania (5) vol. 15 (1925).

(7) G. Aprile, Giornale di Matematica vol. 70 (1932) pp. 196–216.

(8) After the present paper had been sent to the editor, I received from Professor Virgil Snyder an offprint of his *Cremona involutions belonging to the Bordiga surface in* $[4]$, Revista de Tucumán (A) vol. 2 (1941) pp. 203–210, in which he describes a line congruence of order one in $[n]$ consisting of the $(n-1)$-secants of the generalized Bordiga. This is type $(n-1)_n$ in my classification.

in $[n]$ there is a V^k_{n-k-1}, all of whose points are fundamental. (V^k_0 means k points.) This variety may be irreducible or composite. In the latter case the dimension of each component is $n-k-1$ while their orders may be any positive integers $A_1, A_2, \cdots, A_w$, whose sum is k.

The *type symbol* $(A_1A_2A_3 \cdots A_w)_n$ will be used to indicate an $[n-k]$-congruence of order one in $[n]$ whose generic $[n-k]$ has on it w distinct fundamental varieties of orders $A_1, A_2, \cdots, A_w$, respectively. When the meaning of the type symbol becomes completely evident (§4), it will be seen that permutation of the A_i in the type symbol changes the type. Therefore *there are 2^{k-1} types of $[n-k]$-congruences of order one in $[n]$, for which the fundamental loci on a generic $[n-k]$ are all distinct.*

Since some or all of the fundamental varieties on a generic $[n-k]$ of the congruence may be coincident (§6), we divide $[n-k]$-congruences of order one in $[n]$ into those for which these components are all distinct and those for which some or all are coincident. Until §6, congruences in the former category will occupy us.

2. **Fixed base.** The equations

$$\begin{aligned} &\lambda_1 x_{11} + \cdots + \lambda_{r+1} x_{1\,r+1} = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\lambda_1 x_{r1} + \cdots + \lambda_{r+1} x_{r\,r+1} = 0, \end{aligned} \tag{2.1}$$

in which the λ_j are parameters and the x_{ij} are linear forms in the homogeneous projective point coordinates $x_0, x_1, \cdots, x_n$ of $[n]$, having arbitrary constant coefficients, represent an ∞^r-system of $[n-r]$'s in $[n]$.

A necessary and sufficient condition for an $[n-r]$ of this system to pass through an arbitrary point P of $[n]$ is

$$\begin{aligned} &\lambda_1 p_{11} + \cdots + \lambda_{r+1} p_{1\,r+1} = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\lambda_1 p_{r1} + \cdots + \lambda_{r+1} p_{r\,r+1} = 0, \end{aligned} \tag{2.2}$$

where p_{ij} is the result of substituting the coordinates of P in x_{ij}.

The symbol

$$\left\| \begin{matrix} x_{11} \cdots x_{1\,r+1} \\ \cdots \quad \cdots \\ x_{r1} \cdots x_{r\,r+1} \end{matrix} \right\| = 0 \tag{2.3}$$

means $r+1$ simultaneous equations $X_h = 0$ ($h = 1, 2, \cdots, r+1$), where X_h is $(-1)^{h+1}$ times the determinant formed from the matrix in (2.3) by omitting the hth column.

Denoting by P_h the result of substituting p_{ij} for x_{ij} in X_h, we have from (2.2)

$$\lambda_1:\lambda_2:\cdots:\lambda_{r+1} = P_1:P_2:\cdots:P_{r+1}.$$

Therefore

$$\begin{aligned} &P_1x_{11}+\cdots+P_{r+1}x_{1\,r+1}=0,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ &P_1x_{r1}+\cdots+P_{r+1}x_{r\,r+1}=0, \end{aligned} \tag{2.4}$$

are the equations of the $[n-r]$ of system (2.1) through point P.

Now a necessary condition for any other point Q to lie on (2.4), the $[n-r]$ through P, is

$$\begin{aligned} &P_1q_{11}+\cdots+P_{r+1}q_{1\,r+1}=0,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ &P_1q_{r1}+\cdots+P_{r+1}q_{r\,r+1}=0, \end{aligned} \tag{2.5}$$

where the q_{ij} are obtained by substituting the coordinates of Q in the x_{ij}. From which $P_1:P_2:\cdots:P_{r+1}=Q_1:Q_2:\cdots:Q_{r+1}$. Thus every point on (2.4) determines that same $[n-r]$, and the ratios $P_1:P_2:\cdots:P_{r+1}$ are unique for the $[n-r]$ of the system (2.1) passing through P.

Since (2.4) defines one $[n-r]$ through an arbitrary point P of $[n]$ and every point of (2.4) determines that same $[n-r]$, it follows that *the totality of $[n-r]$'s given by* (2.4) *for all points P of $[n]$ form an $[n-r]$-congruence of order one in $[n]$.*

(2.3) is a determinantal locus(9). Since it will appear that (2.3) serves as a carrier or base for flat space congruences of order one, (2.3) will be called a *fixed base.*

Indicate the fixed base (2.3) by $\mathcal{A}_1$. It is a variety of dimension $n-2$ and order $C_{r+1,2}$(10). Degenerate cases, for which the order of the fixed base is lower, will be treated in a separate note.

The points of $\mathcal{A}_1$ are fundamental points for the $[n-r]$-congruence of order one in $[n]$ whose generic $[n-r]$ is given by (2.4).

For a general point P not on $\mathcal{A}_1$, at least one of the P_j $(j=1, 2, \cdots, r+1)$ is different from zero. For convenience, let $P_{r+1}\neq 0$. If all the other P_j are zero, it is obvious that the intersection of the $[n-r]$ through P, given by (2.4), with the fixed base (2.3), is the same as the intersection of (2.4) with

$$X_{r+1} \equiv \begin{vmatrix} x_{11}\cdots x_{1r} \\ \cdots\cdots \\ x_{r1}\cdots x_{rr} \end{vmatrix} = 0. \tag{2.6}$$

If some other P_j, say P_i, is also different from zero, (2.6) can be rewritten

(9) T. G. Room, *The geometry of determinantal loci*, Cambridge, 1938, p. 33.

(10) Room, loc. cit. pp. 34, 43.

$$\frac{1}{P_i}\begin{vmatrix} x_{11}\cdots x_{1\,i-1}(P_1x_{11}+\cdots+P_rx_{1r})x_{1\,i+1}\cdots x_{1r} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{r1}\cdots x_{r\,i-1}(P_1x_{r1}+\cdots+P_rx_{rr})x_{r\,i+1}\cdots x_{rr} \end{vmatrix}=0.$$

Using (2.4), this becomes

$$\frac{1}{P_i}\begin{vmatrix} x_{11}\cdots x_{1\,i-1}(P_{r+1}x_{1\,r+1})x_{1\,i+1}\cdots x_{1r} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{r1}\cdots x_{r\,i-1}(P_{r+1}x_{r\,r+1})x_{r\,i+1}\cdots x_{rr} \end{vmatrix}=0,$$

which is $P_{r+1}X_i/P_i=0$.

Since the selection of P_{r+1} as the nonzero P_j was arbitrary, it follows that *the intersection of the* $[n-r]$ *through* P, *given by* (2.4), *with the fixed base* (2.3), *is the same as the intersection of* (2.4) *with any* $X_j=0$ *for which* $P_j\neq 0$.

This intersection is a determinantal primal in the space $[n-r]$ through P and is of order r. Its equation is formed by imposing the r linear conditions (2.4) on the running coordinates $x_0, x_1, \cdots, x_n$ in each element x_{ij} of $X_j=0$.

In the fixed base (2.3), let $x_{ij}\equiv\sum_{k=0}^{n}a_{ijk}x_i$, where a_{ijk} are constants. Equations (2.1) may be rewritten

$$\begin{aligned} &\lambda_{10}x_0+\cdots+\lambda_{1n}x_n=0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\lambda_{r0}x_0+\cdots+\lambda_{rn}x_n=0, \end{aligned} \tag{2.7}$$

where $\lambda_{ik}\equiv\sum_{j=1}^{r+1}\lambda_j a_{ijk}$. In general, for each set of values $\lambda_1, \cdots, \lambda_{r+1}$, these equations are independent and represent an $[n-r]$ of the congruence based on (2.3). A necessary and sufficient condition for equations (2.7) to be dependent is that the rank of the matrix

$$\left\|\begin{matrix} \lambda_{10}\cdots\lambda_{1n} \\ \cdots\cdots\cdots \\ \lambda_{r0}\cdots\lambda_{rn} \end{matrix}\right\| \tag{2.8}$$

be m $(m<r)$. In such event, (2.8) represents a variety Λ of dimension[11] $r-m$ $(n+1-r+m)$ in a space $[r]$ whose generic point has coordinates $(\lambda_1, \cdots, \lambda_{r+1})$. The order of Λ is given by a formula to Segre[12]. To each point of Λ corresponds a set of values $\lambda_1, \cdots, \lambda_{r+1}$ which determine in (2.1), not an $[n-r]$, but an $[n-m]$.

For example, the type $(3)_4$ line congruence of order one in $[4]$ consists of the trisecants of a Bordiga sextic surface. This surface is the fixed base $\mathcal{A}_1$, all of whose points are fundamental. But there are 10 planes in $[4]$, each of which cuts the Bordiga in a cubic curve. Through a general point of such a

(11) Room, loc. cit. p. 36.

(12) Room, loc. cit. p. 42.

plane a pencil of lines can be drawn, each of which is a trisecant of the Bordiga surface. The equations of the 10 planes are obtained by substituting in (2.1) values of $\lambda_1, \cdots, \lambda_{r+1}$, which make the rank of matrix (2.8) be $m=2$.

Let $t_{i_1 i_2 \cdots i_r} = t_{(i)}$ be the r-row minors of the matrix (2.8). These minors $t_{(i)}$ are forms in $\lambda_1, \cdots, \lambda_{r+1}$. *An* $[n-r]$ *with coordinates* $\bar{t}_{i_1 i_2 \cdots i_r} = \bar{t}_{(i)}$ *belongs to the congruence if and only if its coordinates* $\bar{t}_{(i)}$ *satisfy all homogeneous algebraic relations which are satisfied identically by the* $t_{(i)}$; *that is, if and only if* $F(\cdots, t_{i_1 i_2 \cdots i_r}, \cdots) = 0$ (*identically in* $\lambda_1, \lambda_2, \cdots, \lambda_{r+1}$) *implies* $F(\cdots, \bar{t}_{i_1 i_2 \cdots i_r}, \cdots) = 0$.

3. **Dependent base.** When $r=k$ in §2, there results an $[n-k]$-congruence of order one in $[n]$ for which the fundamental variety on a generic $[n-k]$ is irreducible.

When $r<k$, other $[n-k]$-congruences of order one may be constructed as follows. For these, the fundamental variety on a generic $[n-k]$ is composite.

Consider

$$\left\| \begin{matrix} x_{11}^{(2)} & \cdots & x_{1\,s+1}^{(2)} \\ \cdot & \cdots & \cdot \\ x_{s1}^{(2)} & \cdots & x_{s\,s+1}^{(2)} \end{matrix} \right\| = 0, \tag{3.1}$$

where $x_{ij}^{(2)} \equiv \sum_{k=0}^{n} a_{ijk}^{(2)} x_k$, in which $a_{ijk}^{(2)}$ are forms of order σ_{21} (any positive integer or zero) in the coordinates P_h ($h=1, 2, \cdots, r+1$) of (2.4).

(3.1) will be called a *dependent base.* Since this concept seems to be new, it is emphasized that the $x_{ij}^{(2)}$ are linear forms in the projective point coordinates of $[n]$, *having coefficients which are forms of order* σ_{21} *in the* $P_1, P_2, \cdots, P_{r+1}$ *of the* $[n-r]$ *through arbitrary* P.

Geometrically, an arbitrary point P of $[n]$ determines the $[n-r]$ of (2.4) through it whose $P_1, \cdots, P_{r+1}$ make (3.1) a fixed base with elements that are linear forms in the running coordinates $x_0, \cdots, x_n$, with constant coefficients determined by $P_1, \cdots, P_{r+1}$. Thus there is defined in the $[n-r]$ through P an $[n-r-s]$-congruence of order one whose $[n-r-s]$ through P is given by equations (2.4) along with the equations

$$\begin{aligned} &P_1^{(2)} x_{11}^{(2)} + \cdots + P_{s+1}^{(2)} x_{1\,s+1}^{(2)} = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &P_1^{(2)} x_{s1}^{(2)} + \cdots + P_{s+1}^{(2)} x_{s\,s+1}^{(2)} = 0, \end{aligned} \tag{3.2}$$

in which $P_j^{(2)}$ is obtained by substituting the coordinates of P in $X_j^{(2)}$, and $X_j^{(2)}$ is $(-1)^{j+1}$ times the determinant formed from the matrix of (3.1) by omitting the jth column.

The ratios $P_1^{(2)} : P_2^{(2)} : \cdots : P_{s+1}^{(2)}$ are characteristic of the $[n-s]$ through P, given by (3.2).

If $r+s=k$, equations (2.4) and (3.2) together define a unique $[n-k]$ through a generic point P of $[n]$, and the totality of these $[n-k]$'s constitute an $[n-k]$-congruence of order one in $[n]$. The fundamental variety on each $[n-k]$ of the congruence is composite, consisting of a V^r_{n-k-1} in which the $[n-k]$ intersects (2.3) and a V^s_{n-k-1} in which the $[n-k]$ intersects (3.2).

If $r+s<k$, another dependent base may be constructed whose elements are linear forms in $x_0, x_1, \cdots, x_n$ with coefficients that are homogeneous of order σ_{31} in the $P_1, \cdots, P_{r+1}$ of (2.4) and also homogeneous of order σ_{32} in the coordinates $P^{(2)}_1, \cdots, P^{(2)}_{s+1}$ of (3.2) (σ_{31} and σ_{32} any positive integers or zero).

This process may be continued until the number of equations defining the generic flat space of the congruence is equal to k.

The effect of these dependent bases is to associate with each element of a flat space congruence of order one in $[n]$, in the most general possible manner, a unique congruence within that element.

4. **Type** $(A_1\ A_2 \cdots A_w)_n$. We are now prepared to describe completely the general type $(A_1A_2 \cdots A_w)_n$, where the A_i are any positive integers whose sum is $k<n$.

Associated with the positive integer $A_1=r$ in first position in the type symbol is the fixed base

$$\left\| \begin{matrix} x_{11} \cdots x_{1\,r+1} \\ \cdot \quad \cdots \quad \cdot \\ x_{r1} \cdots x_{r\,r+1} \end{matrix} \right\| = 0, \tag{4.1}$$

in which the x_{ij} are linear forms in $x_0, x_1, \cdots, x_n$ having arbitrary constant coefficients.

Then (§2),

$$\begin{aligned} &P_1x_{11} + \cdots + P_{r+1}x_{1\,r+1} = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &P_1x_{r1} + \cdots + P_{r+1}x_{r\,r+1} = 0, \end{aligned} \tag{4.2}$$

where the P_i are obtained by substituting the coordinates of an arbitrary point P of $[n]$ in X_i, and X_i is $(-1)^{i+1}$ times the determinant formed from the matrix of (4.1) by omitting the ith column, represent an $[n-r]$ through P.

Associated with the positive integer $A_h=s$ in hth position in the type symbol ($h=2, 3, \cdots, w-1$) is the dependent base

$$\left\| \begin{matrix} x^{(h)}_{11} \cdots x^{(h)}_{1\,s+1} \\ \cdot \quad \cdots \quad \cdot \\ x^{(h)}_{s1} \cdots x^{(h)}_{s\,s+1} \end{matrix} \right\| = 0, \tag{4.3}$$

where $x^{(h)}_{ij} \equiv \sum_{k=0}^{n} a^{(h)}_{ijk}x_k$, in which $a^{(h)}_{ijk}$ are homogeneous of order σ_{h1} in the

$P_1, \cdots, P_{r+1}$ of the flat space (4.2) through P, based on the fixed base (4.1) associated with A_1. The $a_{ijk}^{(h)}$ are also homogeneous of order σ_{hm} in the $P_1^{(m)}, P_2^{(m)}, \cdots$ of the flat space through P based on the dependent base associated with A_m $(m=2, 3, \cdots, h-1)$. The σ_{h1} and σ_{hm} are any positive integers or zero.

Then

$$
\begin{aligned}
&P_1^{(h)} x_{11}^{(h)} + \cdots + P_{s+1}^{(h)} x_{1\,s+1}^{(h)} = 0,\\
&\cdots\cdots\cdots

such $[n-k]$'s, for all points P of $[n]$, form an $[n-k]$-congruence of order one in $[n]$, whose type symbol is $(A_1A_2 \cdots A_w)_n$.

5. **Fundamental points.** For the $[n-k]$-congruence of order one in $[n]$ whose type symbol is $(A_1A_2 \cdots A_w)_n$, there are w varieties $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_w$, each of dimension $n-2$, all of whose points are fundamental. Through each point of any $\mathcal{A}_i$ there pass infinitely many $[n-k]$'s.

$\mathcal{A}_1$ is the fixed base (4.1). We have seen (§2) that $\mathcal{A}_1$ is a variety of dimension $n-2$ and order $\alpha_1 = C_{r+1,2}$.

The dependent base associated with A_2 of the type symbol is

$$\left\| \begin{matrix} x_{11}^{(2)} & \cdots & x_{1\,s+1}^{(2)} \\ \cdot & \cdots & \cdot \\ x_{s1}^{(2)} & \cdots & x_{s\,s+1}^{(2)} \end{matrix} \right\| = 0 \tag{5.1}$$

where $x_{ij}^{(2)}$ are linear forms in $x_0, \cdots, x_n$, whose coefficients are forms of order σ_{21} in $P_1, \cdots, P_{r+1}$.

Denote by $\mathbf{x}_{ij}^{(2)}$ and $\mathbf{F}_j^{(2)}$ the result of changing to $x_0, \cdots, x_n$ the coordinates of P wherever they appear in $x_{ij}^{(2)}$ and $F_j^{(2)}$, respectively. Now

$$\left\| \begin{matrix} \mathbf{x}_{11}^{(2)} & \cdots & \mathbf{x}_{1\,s+1}^{(2)} \\ \cdot & \cdots & \cdot \\ \mathbf{x}_{s1}^{(2)} & \cdots & \mathbf{x}_{s\,s+1}^{(2)} \end{matrix} \right\| = 0, \tag{5.2}$$

is a variety of dimension $n-2$. Its order is the number of points in which it intersects a general plane of $[n]$. Since (5.2) has no special position relative to the frame of reference, its order will be given by the number of points in which it intersects the particular plane π, whose equations are $x_3 = x_4 = \cdots = x_n = 0$. Since each element $\mathbf{x}_{ij}^{(2)}$ of the matrix (5.2) is homogeneous of degree σ_{21} in $F_1, \cdots, F_{r+1}$, every point of $\mathcal{A}_1$ lies on (5.2). That is, (5.2) is a composite variety of dimension $n-2$, consisting of $\mathcal{A}_1$ and a residual which is $\mathcal{A}_2$. We seek the order of $\mathcal{A}_2$. The order of $\mathcal{A}_2$ is the number of points, not on $\mathcal{A}_1$, in which the plane π intersects (5.2).

Denote by $\xi_{ij}^{(2)}$, Φ_j, and $\Phi_j^{(2)}$, the result obtained by equating to zero $x_3, x_4, \cdots, x_n$ in $\mathbf{x}_{ij}^{(2)}$, F_j, and $\mathbf{F}_j^{(2)}$, respectively. Salmon([13]) has pointed out that the locus

$$\left\| \begin{matrix} \xi_{11}^{(2)} & \cdots & \xi_{1\,s+1}^{(2)} \\ \cdot & \cdots & \cdot \\ \xi_{s1}^{(2)} & \cdots & \xi_{s\,s+1}^{(2)} \end{matrix} \right\| = 0 \tag{5.3}$$

([13]) G. Salmon, *Modern higher algebra*, 4th edition, Dublin, 1885, p. 287.

is the intersection of $\Phi_1^{(2)}=0$ and $\Phi_2^{(2)}=0$ after rejecting those points of the locus

$$\left\| \begin{matrix} \xi_{13}^{(2)} & \cdots & \xi_{s3}^{(2)} \\ \cdot & \cdots & \cdot \\ \xi_{1\,s+1}^{(2)} & \cdots & \xi_{s\,s+1}^{(2)} \end{matrix} \right\| = 0$$

which do not lie on $\Phi_1^{(2)}=0$ and $\Phi_2^{(2)}=0$. (Notice that the rows of this latter matrix are the columns of (5.3) common to $\Phi_1^{(2)}$ and $\Phi_2^{(2)}$.)

Let $s=1$ in (5.3). Then $\Phi_1^{(2)}=0$ is a plane curve of order $r\sigma_{21}+1$. Since $\Phi_1^{(2)}$ is homogeneous of degree σ_{21} in $\Phi_1, \Phi_2, \cdots, \Phi_{r+1}$, and $\Phi_1=\Phi_2=\cdots=\Phi_{r+1}=0$ is $C_{r+1,2}$ points, each of these points is σ_{21}-fold on the plane curve $\Phi_1^{(2)}=0$. Similarly, $\Phi_2^{(2)}=0$ is a plane curve of order $r\sigma_{21}+1$ on which each of the $C_{r+1,2}$ points, $\Phi_1=\Phi_2=\cdots=\Phi_{r+1}=0$, is σ_{21}-fold. Therefore $\Phi_1^{(2)}=0$ and $\Phi_2^{(2)}=0$ intersect in $(r\sigma_{21}+1)^2$ points, among which the $C_{r+1,2}$ points count $C_{r+1,2}\sigma_{21}^2$. That is, when $s=1$, the number of points, not on $\mathcal{A}_1$, in which (5.1) intersects π is $(r\sigma_{21}+1)^2-C_{r+1,2}\sigma_{21}^2$. This is the order of $\mathcal{A}_2$ when $s=1$.

Now consider the case $s=2$ in (5.3). $\Phi_1^{(2)}=0$ is a plane curve of order $2(r\sigma_{21}+1)$. Since $\Phi_1^{(2)}$ is homogeneous of degree $2\sigma_{21}$ in $\Phi_1, \cdots, \Phi_{r+1}$, each of the $C_{r+1,2}$ points, $\Phi_1=\Phi_2=\cdots=\Phi_{r+1}=0$, is $2\sigma_{21}$-fold on the plane curve $\Phi_1^{(2)}=0$. Similarly, $\Phi_2^{(2)}=0$ is a plane curve of order $2(r\sigma_{21}+1)$, on which each of the $C_{r+1,2}$ points is $2\sigma_{21}$-fold. $\Phi_1^{(2)}=0$ and $\Phi_2^{(2)}=0$ intersect in $4(r\sigma_{21}+1)^2$ points in which the $C_{r+1,2}$ points count as $4\sigma_{21}^2C_{r+1,2}$. That is, $\Phi_1^{(2)}=0$ and $\Phi_2^{(2)}=0$ intersect in $4(r\sigma_{21}+1)^2-4\sigma_{21}^2C_{r+1,2}$ points not on $\mathcal{A}_1$. We have already seen ($s=1$) that

$$\left\| \xi_{13}^{(2)} \quad \xi_{23}^{(2)} \right\| = 0$$

consists of $(r\sigma_{21}+1)^2-\sigma_{21}^2C_{r+1,2}$ points not on $\mathcal{A}_1$. Therefore, for $s=2$, (5.3) consists of $3\{(r\sigma_{21}+1)^2-C_{r+1,2}\sigma_{21}^2\}$ points not on $\mathcal{A}_1$. This is the order of $\mathcal{A}_2$ for $s=2$.

By induction it can be shown that the order of $\mathcal{A}_2$, for s any positive integer, is $\alpha_2 = C_{s+1,2}\{(r\sigma_{21}+1)^2 - C_{r+1,2}\sigma_{21}^2\}$.

The determination of the orders of $\mathcal{A}_3, \mathcal{A}_4, \cdots, \mathcal{A}_w$ becomes increasingly tedious and does not differ essentially from the above.

When $\sigma_{hk}\neq 0$ in the dependent base (4.3), the $P_1^{(k)}, P_2^{(k)}, \cdots$ appear homogeneously in the coefficients of the linear forms $x_{ij}^{(h)}$, and this causes $\mathcal{A}_h$ to intersect $\mathcal{A}_k$ in a variety of dimension $n-3$ ($k<h$). If, moreover, $\sigma_{kg}\neq 0$ ($g<k$), $\mathcal{A}_h$ intersects $\mathcal{A}_g$ in a variety of dimension $n-3$. But $P_1^{(g)}, P_2^{(g)}, \cdots$ appear in the $P_1^{(h)}, P_2^{(h)}, \cdots$, and therefore also in the coefficients of the linear forms $x_{ij}^{(h)}$, and consequently $\mathcal{A}_h$ intersects $\mathcal{A}_g$ in a variety of dimension $n-3$. *If $\mathcal{A}_k$ intersects $\mathcal{A}_g$ in a variety of dimension $n-3$ and $\mathcal{A}_h$ intersects $\mathcal{A}_k$ in a variety of dimension $n-3$ ($g<k<h$), then $\mathcal{A}_h$ intersects $\mathcal{A}_g$ in a variety of dimension $n-3$.* The three varieties of dimension $n-3$, just mentioned, are

distinct. Subject only to this restriction, we cause $\mathcal{A}_h$ to intersect $\mathcal{A}_g$ in a V_{n-3} or not, according as we choose σ_{hg} different from, or equal to, zero in constructing our flat space congruence of type $(A_1 \cdots A_w)_n$.

6. **Coincident fundamental points.** For convenience in the type symbol $(A_1 A_2 \cdots A_{h-1} A_h \cdots A_w)_n$, let $A_1+A_2+\cdots+A_{h-1}=u$. A unique $[n-u]$ of the congruence $(A_1A_2 \cdots A_{h-1})_n$ passes through P. This $[n-u]$ intersects $\mathcal{A}_i$ in $\mathcal{D}_i$ $(i<h)$. $\mathcal{D}_i$ is a determinantal primal of space $[n-u]$ and is of dimension $n-u-1$ and order A_i (§2).

By means of the dependent base (4.3) there is determined in $[n-u]$ a fixed base D, whose matrix contains s rows.

When $s=A_i$ $(i<h)$, D can clearly lie on $\mathcal{A}_i$. When $s=A_i-1$, there are two systems of fixed bases, whose matrices have s rows, on $\mathcal{D}_i$[14] and thus D can lie on $\mathcal{D}_i$.

Therefore, *whenever in the type symbol an integer is equal to, or one less than, an integer further to the left, there exists an* $[n-k]$*-congruence of order one in* $[n]$ *with coincidences among the fundamental points.*

In the case of a line congruence having $A_w=1$ in its type symbol, the plane through P of the associated plane congruence $(A_1A_2 \cdots A_{w-1})_n$ intersects $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_{w-1}$ in plane curves $\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_{w-1}$, respectively, and $\mathcal{A}_w$ in a single point. This point can be made to lie on any of the plane curves, and therefore *for the type symbol* $(A_1A_2 \cdots A_{w-1}1)_n$, *with* $A_1+A_2+\cdots+A_{w-1}+1=n-1$, *there exist line congruences of order one in* $[n]$ *whose fundamental points associated with* A_w *are coincident with those associated with any earlier* A_i.

7. **Ordinary space.** In the interest of exposition, the foregoing will now be applied to the well known situation in three-dimensional space.

There is only one plane congruence[15] of order one in [3]. Its type symbol is $(1)_3$ and its fixed base is

$$\|x_{11} \quad x_{12}\| = 0, \tag{7.1}$$

where x_{1j} $(j=1, 2)$ are linear forms in x_0, x_1, x_2, x_3, with arbitrary constant coefficients.

(7.1) represents a fixed line in [3]. The unique plane through this line and a generic point P is

$$P_1x_{11} + P_2x_{12} = 0, \tag{7.2}$$

where P_1 is the result of substituting the coordinates of P in x_{12} and P_2 is the result of substituting the coordinates of P in $-x_{11}$. The totality of planes (7.2) form a plane congruence of order one in [3].

The locus of fundamental points, $\mathcal{A}_1$, is the line (7.1).

---

[14] Room, loc. cit. p. 106.

[15] While a pencil of planes in [3] is not ordinarily called a congruence, it fulfills our definition $(k=1)$, and such usage makes for uniformity of language.

In [3] there are line congruences of order one with type symbols $(2)_3$ and $(11)_3$. Each gives rise to an additional line congruence having coincident fundamental points.

The fixed base for type $(2)_3$ is

$$\left\| \begin{matrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{matrix} \right\| = 0, \tag{7.3}$$

where the x_{ij} are linear forms in x_0, x_1, x_2, x_3, having arbitrary constant coefficients. (7.3) is a twisted cubic curve whose unique bisecant through a generic point P of [3] is

$$\begin{aligned} P_1x_{11} + P_2x_{12} + P_3x_{13} &= 0, \\ P_1x_{21} + P_2x_{22} + P_3x_{23} &= 0. \end{aligned} \tag{7.4}$$

The P_j are obtained by substituting the coordinates of P in X_j $(j=1, 2, 3)$, and X_j is $(-1)^{j+1}$ times the determinant formed from the matrix of (7.3) by omitting the jth column.

The totality of rays (7.4) is a line congruence of order one in [3] with type symbol $(2)_3$ and whose irreducible locus of fundamental points $\mathcal{A}_1$ is a twisted cubic curve (7.3).

If all the planes $x_{ij}=0$ $(i=1, 2; j=1, 2, 3)$ pass through the same point O, the twisted cubic (7.3) degenerates into three lines through O. The congruence now consists of the bundle of lines through O. The three lines (7.3) through O are not fundamental but are members of the congruence. The only fundamental point is the vertex O. This oldest line congruence of order one in [3] is thus a special case of our type $(2)_3$.

In the line congruence of type $(11)_3$, the fixed base is the line (7.1). The terminal base is given by

$$\| x_{11}^{(2)} \quad x_{12}^{(2)} \| = 0, \tag{7.5}$$

where $x_{ij}^{(2)} \equiv \sum_{k=0}^{3} a_{ijk}^{(2)} x_k$, and $a_{ijk}^{(2)}$ are forms of order τ_{j1} (any positive integer or zero) in P_1, P_2 of plane (7.2). For each plane (7.2), (7.5) represents a line in [3] intersecting that plane in a point O.

The equation

$$P_1^{(2)} x_{11}^{(2)} + P_2^{(2)} x_{12}^{(2)} = 0, \tag{7.6}$$

in which $P_1^{(2)}$ and $P_2^{(2)}$ are obtained by substituting the coordinates of P for the running coordinates in $x_{12}^{(2)}$ and $-x_{11}^{(2)}$, respectively, represents for each $P_1 : P_2$ a plane through P.

Equations (7.2) and (7.6) together represent the ray through P, lying in plane (7.2) and through point O. The totality of rays given by (7.2) and (7.6) form a line congruence of order one in [3].

The locus of fundamental points is composite, consisting of $\mathcal{A}_1$, the fixed

line (7.1), and $\mathcal{A}_2$, a space curve of order $\tau_{11}+\tau_{21}+1$. The equations of $\mathcal{A}_2$ are formed from (7.5) by replacing the coordinates of P in the $a^{(2)}_{1jk}$ by the running coordinates x_0, x_1, x_2, x_3. These equations contain the line (7.1) counted $\tau_{11}\tau_{21}$ times. The space curve has $\tau_{11}+\tau_{21}$ intersections with line $\mathcal{A}_1$. Point O is the intersection of curve $\mathcal{A}_2$ with plane (7.2), not on $\mathcal{A}_1$.

If, in the preceding type $(11)_3$ congruence, we impose the restriction that the determinant of the coefficients of x_{11}, x_{12}, $x^{(2)}_{11}$, and $x^{(2)}_{12}$ shall vanish for all values of $P_1:P_2$, the two fundamental points on a generic ray of the congruence will coincide on the line (7.1). The locus of fundamental points is the line (7.1), provided that not both τ_{11} and τ_{21} are zero. There is established a $(\tau_{11}+\tau_{21}, 1)$ correspondence between the planes and points on (7.1). A generic point P of $[3]$ determines a plane through line (7.1) and this plane in turn determines one point O on (7.1). Line PO is the unique ray of the congruence through P.

It may be of interest to observe that while it has long been recognized that this latter congruence is a limiting case of the line congruence of type $(11)_3$, it has hitherto been thought impossible to obtain the coincident case by specialization of the parameters in the equations of type $(11)_3$.

8. **Aprile's congruence.** The $[n-2]$'s determined by the point groups of a g^{n-1}_{n-1} belonging to a rational normal curve of order $2n-3$ in $[2n-3]$ form an $[n-2]$-congruence of order one in $[2n-3]$. The intersections of the $[n-2]$'s of this congruence with an $[n]$ of $[2n-3]$ are lines of a line congruence of order one in $[n]$.

This construction is due to Aprile([16]). We shall show that it is a restricted case of our type $(n-1)_n$ congruence.

Let the parametric equations of the normal rational curve C in $[2n-3]$ be $x_i=\lambda^{2n-3-i}\mu^i$ $(i=0, 1, 2, \cdots, 2n-3)$.

Consider the fixed base

$$\left\| \begin{matrix} x_0 & x_1 & x_2 \cdots x_{n-1} \\ x_1 & x_2 & x_3 \cdots x_n \\ \cdot & \cdot & \cdots \cdot \\ x_{n-2} & x_{n-1} & x_n \cdots x_{2n-3} \end{matrix} \right\| = 0, \tag{8.1}$$

whose elements $x_0, x_1, \cdots, x_{2n-3}$ are point coordinates in $[2n-3]$. Then

$$\begin{aligned} P_1x_0 &+ P_2x_1 &+ \cdots + P_nx_{n-1} &= 0, \\ P_1x_1 &+ P_2x_2 &+ \cdots + P_nx_n &= 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ P_1x_{n-2} &+ P_2x_{n-1} &+ \cdots + P_nx_{2n-3} &= 0, \end{aligned} \tag{8.2}$$

([16]) Aprile, loc. cit. p. 204.

are the equations of a unique $[n-2]$ through a generic point P of $[2n-3]$.

(8.2) intersects the normal curve C in the $n-1$ points whose parameters are given by

$$P_1\lambda^{n-1} + P_2\lambda^{n-2}\mu + P_3\lambda^{n-3}\mu^2 + \cdots + P_n\mu^{n-1} = 0,$$

as may be seen by substituting the parametric equations of curve C in (8.2).

Thus Aprile's $[n-2]$-congruence in $[2n-3]$ is the same as that given by (8.2), and the section of this congruence by an $[n]$ of $[2n-3]$ is Aprile's line congruence of order one in $[n]$.

The $[n]$-section of (8.1) is a fixed base. But (8.1) is what Room calls a partly symmetric determinantal locus, as is also its $[n]$-section. Hence[17], the freedom of the fixed base of Aprile's line congruence of order one in $[n]$ is

$$\delta = (1/2)(n-1)(n-2)(n+1) - (n-1)n + 1$$

less than that of our fixed base, for $n>3$.

---

[17] Room, loc. cit. p. 160.

UNIVERSITY OF ARIZONA,
TUCSON, ARIZ.

ON THE APPROXIMATION TO INTEGRABLE FUNCTIONS BY INTEGRAL FUNCTIONS

BY

H. KOBER

1. **Introduction.** The problem of uniform approximation to a function $f(z)$ on some finite point set D of the z-plane ($z=x+iy$) (a) by polynomials and (b) by rational functions is classical. The theory has been summarized by J. L. Walsh(1), who himself has proved a great number of important theorems some of which are fundamental.

Let D be a point set containing the point $z=\infty$. In that case only problem (b) has a meaning; there is, however, another generalization of problem (a), the approximation to an arbitrary function, defined in D, by integral functions. Not much attention appears to have been paid to that problem up to now, except for the work of S. Bernstein(2).

Among the point sets D extending to infinity the following are the simplest ones and appear to be of particular interest:

I. The real axis $-\infty<x<\infty$, $y=0$.

II. The positive part of the real axis $0<x<\infty$, $y=0$.

III. The half-plane $-\infty<x<\infty$, $y>0$.

IV. The strip $-\infty<x<\infty$, $a<y<b$.

V. The exterior of a circle.

In case V, however, the uniform approximation by integral functions is not a reasonable problem, nor for any other domain of which $z=\infty$ is an interior point. In that case there is a circle C, with radius R and center at the origin, such that the boundary of D lies interior to C and C interior to D. Let $\{f_n(z)\}$ $(n=1, 2, \cdots)$ be a sequence of integral functions approximating to $f(z)$, then $|f(z)-f_n(z)|\leqq\epsilon_n$ for $|z|\geqq R$, where $\epsilon_n\to 0$ as $n\to\infty$. Hence $|f_n(z)-f_1(z)|\leqq\epsilon_1+\epsilon_n$, uniformly for $|z|\geqq R$, and, by the maximum-modulus theorem, for $|z|<R$. By the Liouville theorem, the function $f_n(z)-f_1(z)$ reduces to a constant, c_n, say, therefore $f(z)$ is an integral function, represented in the form $f(z)=f_1(z)+c+f_n(z)=c-c_n$, where $c=\lim c_n$ $(n\to\infty)$.

Presented to the Society, April 24, 1943; received by the editors July 27, 1942.

(1) *Interpolation and approximation by rational functions in the complex domain*, Amer. Math. Soc. Colloquium Publications, vol. 20, New York, 1935.

(2) *Leçons sur les propriétés extrémales des fonctions analytiques d'une variable réelle*, Paris, 1926. Bernstein gives some results concerning problem I; see Theorem 3(d), §6, of the present paper. See, furthermore, the Bernstein Theorems 12 and 13 (pp. 105–107); they deal with sequences $\{f_n(z)\}$ of integral functions of exponential type, converging uniformly on the real axis to a function $f(z)$ of the same kind, or to a function $f(z)$ possessing some derivatives. A problem connected with the latter one has been treated by M. Krein, C. R. (Doklady) Acad. Sci. URSS. vol. 18 (1938) pp. 619–623.

The cases I and II will be treated in the present paper, the cases III and IV in a further paper. We shall deal with approximation both in the sense of convergence in mean with index p $(0<p<\infty)$ and in the sense of uniform convergence $(p=\infty)$. The order of the approximating integral functions will be required to be finite and to be as small as possible. Whenever in an infinite(3) interval a given function is approximated by integral functions of order ρ then it can also be approximated by integral functions of any order $\rho'>\rho$ (§5). The "best possible" value of ρ is unity in problem I, while it is one-half in problem II (§§6–8). Therefore the integral functions of finite types of the orders $\rho=1$ and $\rho=\frac{1}{2}$ play a dominant part in the results. Problem II presents the greatest difficulties. It will, in substance, be reduced to problem I, for $1<p<\infty$, however, by means of the approximation by rational functions (§7). The problem of "best approximation" will be treated in §9, and a generalization of problem I in §10. A number of auxiliary results will be needed, some of which are possibly of interest in themselves. They will be proved first (§§2–4), in order to simplify the presentation of the main results.

I am greatly indebted to W. H. J. Fuchs, Aberdeen, in particular for suggesting to me the use of the Mittag-Leffler functions.

By $|f(t)|_p$ we denote

$$\left(\int_{-\infty}^{\infty}|f(t)|^p dt\right)^{1/p} \quad \text{or} \quad \operatorname*{ess.\,u.b.}_{-\infty<t<\infty}|f(t)|$$

for $0<p<\infty$ or $p=\infty$, respectively; by $\|f(t)\|_p$ we denote

$$\left(\int_{0}^{\infty}|f(t)|^p dt\right)^{1/p} \quad \text{or} \quad \operatorname*{ess.\,u.b.}_{0<t<\infty}|f(t)|.$$

By $G_\beta^{(\rho)}$ $(0<\beta<\infty)$ we denote the set of integral functions $F(z)$ such that $|F(z)|<A_\epsilon e^{(\beta+\epsilon)|z|^\rho}$ $(A_\epsilon=A(\epsilon))$ for any $\epsilon>0$, by $G^{(\rho)}$ the set of integral functions of order not greater than ρ. We write G_β for $G_\beta^{(1)}$. Evidently $G_\beta^{(\rho)}>G_\gamma^{(\sigma)}$ when $0<\rho<\sigma$.

2. Generalization of a Hardy theorem and of a Plancherel-Pólya theorem.

LEMMA 1. *Let $0<p_j<\infty$, and let $f_j(z)$ be analytic for $|z|\leqq R(j=1,2,\cdots,n)$. Then, for $0<r\leqq R$,*

$$W(r)=\int_0^{2\pi}\sum_{j=1}^{n}|f_j(re^{i\phi})|^{p_j}d\phi \tag{2.1}$$

is a monotone increasing function of r, and $\log W(r)$ is a convex function of $\log r$.

For $n=1$ the result is due to G. H. Hardy(4). The general case can be de-

(3) In consequence of the Weierstrass theorem, in a *finite* interval a continuous function can be approximated uniformly by integral functions of any order.

(4) Proc. London Math. Soc. (2) vol. 14 (1915) pp. 269–277.

duced by an argument due to G. Pólya and G. Szegö([5]).

LEMMA 2. *Let $F(z)\in G_\alpha$ and $F(t)$ be bounded in $(-\infty, \infty)$. Then, for $-\infty<y<\infty$, $|F(x+iy)|\leqq e^{\alpha|y|}|F(t)|_\infty$.*

LEMMA 2′. *Let $P(z)$ be a polynomial, let $F(z)\in G_\alpha$, let $0<p<\infty$ and $P(t)|F(t)|^p\in L_1(-\infty, \infty)$. Then*

(a) $$P(t)|F(t)|^p\to 0 \quad as \quad t\to\pm\infty.$$

(b) $$\int_{-\infty}^{\infty}|P(x+iy)||F(x+iy)|^p dx\leqq e^{\alpha p|y|}\int_{-\infty}^{\infty}|P(t)||F(t)|^p dt.$$

For $0<y<\infty$, Lemma 2 is proved by applying a Phragmén-Lindelöf theorem([6]) to the function $e^{i(\alpha+\epsilon)z}F(z)$ $(\epsilon>0)$ first in both quadrants of the upper half-plane, and then in the upper half-plane, and finally by letting $\epsilon\to 0$. For $P(z)\equiv 1$, Lemma 2′ has been proved by M. Plancherel and G. Pólya([7]); the general case can be deduced in a similar way, starting from the inequality

$$|P(z_0)||\Phi(z_0)|^p\leqq\frac{1}{2\pi}\int_0^{2\pi}|P(z_0+re^{i\phi})||\Phi(z_0+re^{i\phi})|^p d\phi,$$

where $\Phi(z)$ is regular for $|z-z_0|\leqq r$. This inequality is a simple application of Lemma 1.

3. **On the spaces $G_{\alpha,p}$ and $G_{\alpha,p}^{(1/2)}$.** Let p and α be fixed $(0<p\leqq\infty, 0<\alpha<\infty)$. By $G_{\alpha,p}$ or $G_{\alpha,p}^{(1/2)}$ we denote the set of functions $f(z)$, belonging to G_α or $G_\alpha^{(1/2)}$ and such that $f(t)\in L_p(-\infty, \infty)$ or $L_p(0, \infty)$, respectively.

LEMMA 3([8]). *Let $f_n(z)\in G_{\alpha,p}$ $(n=1, 2, \cdots)$ and $|f_m(t)-f_n(t)|_p\to 0$ as $m>n\to\infty$. Then there is an element $f(z)$ of $G_{\alpha,p}$ such that $|f(t)-f_n(t)|_p\to 0$ as $n\to\infty$.*

LEMMA 3′. *The preceding result holds when $G_{\alpha,p}$ is replaced by $G_{\alpha,p}^{(1/2)}$, and $|f_m(t)-f_n(t)|_p$ by $\|f_m(t)-f_n(t)\|_p$, $|f(t)-f_n(t)|_p$ by $\|f(t)-f_n(t)\|_p$.*

The lemmas do not hold whenever the order ρ of the integral functions is greater than 1 or $\frac{1}{2}$, respectively([9]).

([5]) *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925, vol. I, pp. 143, 144, nos. 302, 310. We start from the lemma: Let $f_{j,k}(z)$ $(j=1, 2, \cdots, m; k=1, 2, \cdots, n)$ be regular and $|f_{j,k}(z)|$ be one-valued in the closed domain D, let $0<p_j<\infty$. Then the function $\sum_{k=1}^{n}|f_{1,k}(z)|^{p_1}|f_{2,k}(z)|^{p_2}\cdots|f_{m,k}(z)|^{p_m}$ attains its maximum value on the boundary of D.

([6]) E.g. E. C. Titchmarsh, *The theory of functions*, Oxford, 1932, p. 177.

([7]) Comment. Math. Helv. vol. 10 (1937–1938), §§27–30.

([8]) For $p=\infty$, the result is equivalent to the Bernstein Theorem 13.

([9]) Let $\rho>1$ or $\rho>\frac{1}{2}$, respectively, be fixed, let $f(t)$ be continuous for $-1\leqq t\leqq 1$ and $f(t)=0$ for $|t|\geqq 1$. Then $f(t)$ is not equivalent to an integral function in $(-\infty, \infty)$ nor in $(0, \infty)$. On the other hand, by Theorem 3(b), §6, and by the argument of §5, there are integral functions $g_n(z)$ of the type unity of the order ρ, approximating to $f(t)$ in $(-\infty, \infty)$ or $(0, \infty)$, respectively, as $n\to\infty$; therefore $|g_m(t)-g_n(t)|_p\to 0$ or $\|g_m(t)-g_n(t)\|_p\to 0$. For $p=\infty$ and $\frac{1}{2}<\rho\leqq 1$, how-

Let $f_n(z)\in G_{a,p}$. In consequence of the hypothesis, there is a constant A, independent of n, such that $|f_n(t)|_p<A$, and a function $\Phi(t)\in L_p(-\infty, \infty)$ such that $|\Phi(t)-f_n(t)|_p\to 0$ as $n\to\infty$. Take first $p=\infty$. By Lemma 2, we have $|f_n(z)|<Ae^{a|y|}$,

$$|f_m(z)-f_n(z)|\leqq e^{a|y|}|f_m(t)-f_n(t)|_\infty. \tag{3.1}$$

Consequently the $f_n(z)$ converge to an analytic function $f(z)$, uniformly in any finite domain of the z-plane. Therefore, for any value of z, we have $|f(z)|\leqq Ae^{a|y|}$, and so $f(z)\in G_a$. Since, uniformly in any finite interval $(-M, M)$, $f(t)-f_n(t)\to 0$ as $n\to\infty$, we have $f(t)\equiv\Phi(t)$ in $(-M, M)$ and, therefore, in $(-\infty, \infty)$.

Let now $0<p<\infty$, and put $f_{m,n}=f_m-f_n$. By an argument due to Plancherel and Pólya([10]), we have

$$\begin{aligned}|f_{m,n}(x)|^p &\leqq \frac{1}{\pi}\int_{-1}^{1}d\tau\int_{-1}^{1}|f_{m,n}(x+t+i\tau)|^p dt\\ &\leqq \frac{1}{\pi}\int_{-1}^{1}d\tau\int_{-\infty}^{\infty}|f_{m,n}(s+i\tau)|^p ds.\end{aligned}$$

Using Lemma 2'(b), with $P(t)\equiv 1$, we have

$$|f_{m,n}(x)|\leqq (2/\pi)^{1/p}e^{a}|f_{m,n}(t)|_p \qquad (-\infty<x<\infty). \tag{3.2}$$

By the argument used in the first part of the proof we deduce the existence of a function $f(z)$ belonging to G_a and such that, in any finite interval $(-M, M)$,

$$\int_{-M}^{M}|f(t)-f_n(t)|^p dt\to 0 \quad \text{as} \quad n\to\infty.$$

Hence $f(t)\equiv\Phi(t)$ in $(-M, M)$ and, therefore, in $(-\infty, \infty)$.

The case $f_n(z)\in G_a^{(1/2)}$ is more difficult. Taking $F_n(z)=f_n(z^2)$, $F_n(z)$ is an even function and belongs to G_a. By this transformation, the result for $p=\infty$ is readily deduced from the corresponding result on G_a. Let now $0<p<\infty$, and $F_n(z)=f_n(z^2)$. Then we have

$$\|f_n(t)\|_p=|F_n(t)|\,t|^{1/p}|_p<A \qquad (n=1, 2, \cdots). \tag{3.3}$$

$$\int_0^\infty|f_m(t)-f_n(t)|^p dt=\int_{-\infty}^{\infty}|F_m(t)-F_n(t)|^p|t|\,dt. \tag{3.4}$$

We have to prove the existence of an even function $F(z)\in G_a$ such that $|\{F(t)-F_n(t)\}|t|^{1/p}|_p\to 0$ as $n\to\infty$. In consequence of the hypothesis, there

ever, we can prove: *Let $f_n(z)\in G_a^{(\rho)}$, let $\gamma=2\pi-\pi\rho^{-1}$, let C consist of two straight lines starting at the origin and making the angle γ between them, and let, uniformly for z on C, $f_m(z)-f_n(z)\to 0$ as $m>n\to\infty$. Then there is an element $f(z)$ of $G_a^{(\rho)}$ such that, uniformly on C, $f_n(z)\to f(z)$ as $n\to\infty$.*

([10]) Loc. cit., §30(73). The proof is based on the Hardy theorem.

is an even function $\Phi(t)$ such that $\Phi(t)|t|^{1/p}\in L_p(-\infty, \infty)$ and that $|\phi(t)|t|^{1/p}-F_n(t)|t|^{1/p}|_p\to 0$ as $n\to\infty$. Using the Plancherel-Pólya argument and Lemma 2'(b), we have

$$
\begin{aligned}
|z||F_n(z)|^p &\leqq \frac{1}{\pi}\int_{-1}^{1}d\tau\int_{-1}^{1}|z+t+i\tau||F_n(z+t+i\tau)|^p dt \\
&\leqq \frac{1}{\pi}\int_{-1}^{1}d\tau e^{\alpha p|y+\tau|}\int_{-\infty}^{\infty}|F_n(t)|^p|t|\,dt \leqq \frac{2}{\pi}e^{\alpha p(1+|y|)}\|f_n(t)\|_p^p.
\end{aligned}
\tag{3.5}
$$

Similarly

$$
|z||F_m(z)-F_n(z)|^p \leqq (2/\pi)e^{\alpha p(1+|y|)}\|f_m(t)-f_n(t)\|_p^p. \tag{3.6}
$$

Hence, if k is an integer greater than p^{-1} and M any positive number, we have $z^kF_m(z)-z^kF_n(z)\to 0$ as $m>n\to\infty$, uniformly for $-M\leqq x\leqq M$, $-M\leqq y\leqq M$. Consequently $z^kF_n(z)$ tends to an analytic function $\Psi(z)$ as $n\to\infty$, uniformly in any finite domain of the z-plane. By (3.5) and (3.3), there exists a constant A', independent of n, such that $|z||F_n(z)|^p\leqq A'$ for $|z|\leqq 1$. By the maximum-modulus theorem, we have $|F_n(z)|^p\leqq A'$ for $|z|\leqq 1$. Therefore, for $|z|\leqq 1$, we have $|\Psi(z)|\leqq A'|z|^k$ which shows that $F(z)=z^{-k}\Psi(z)$ is an integral function. Since $z^kF_n(z)\to\Psi(z)$, from (3.5) and (3.3) we deduce that $\Psi(z)$ and, therefore, $F(z)$ belong to G_α. We are left to show that $F(t)\equiv\Phi(t)$ in $(-\infty, \infty)$. Let M be any positive number, and E be the set consisting of the two intervals $(-M, -M^{-1})$ and (M^{-1}, M). By the uniform convergence of the sequence $\{z^kF_n(z)\}$, we have

$$
\int_E |F(t)-F_n(t)|^p|t|\,dt = \int_E |\Psi(t)-t^kF_n(t)|^p|t|^{1-pk}dt \to 0
$$

as $n\to\infty$. Hence $F(t)\equiv\Phi(t)$ in E and, therefore, in $(-\infty, \infty)$, and $F(t)$ is an even function; so is, therefore, $F(z)$.

We remark that, in the terminology of S. Banach, $G_{\alpha,p}$ and $G_{\alpha,p}^{(1/2)}$ are (B) spaces for $1\leqq p\leqq\infty$ [11]. For $0<p<1$, they are (F) spaces if the distance of an element $f(z)$ from the null-element is defined by $|f(t)|_p^p$ or $\|f(t)\|_p^p$, respectively.

4. Further basic lemmas.

LEMMA 4. *Let* $\kappa(z)\in G_1$. *Let* $1\leqq p\leqq\infty$, $1\leqq q\leqq\infty$, $p^{-1}+q^{-1}\geqq 1$ *and* $f(t)\in L_p(-\infty, \infty)$, $\kappa(t)\in L_q(-\infty, \infty)$. *Let*

$$
g_\alpha(z) = \alpha\int_{-\infty}^{\infty}\kappa\{\alpha(t-z)\}f(t)dt \qquad (z=x+iy). \tag{4.1}
$$

Then $g_\alpha(z)$ *belongs to* G_α.

Proof. Let $p'=p/(p-1)$. Then we have $p'\geqq q$ and, by Lemma 2'(a), $\kappa(t)\in L_{p'}(-\infty, \infty)$. By Lemma 2'(b), we have

[11] With "norm" $|f(t)|_p$ or $\|f(t)\|_p$. S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932, Chapters III and IV.

$$(4.2)\qquad \left|\kappa\{\alpha t-\alpha(x\pm iy)\}\right|_{p'}=\left|\kappa(\alpha t-\alpha iy)\right|_{p'}\leqq \alpha^{-1/p'}e^{\alpha|y|}\left|\kappa(t)\right|_{p'}.$$

Therefore, by Hölder's theorem, from (4.1) we deduce that

$$(4.3)\qquad |g_\alpha(z)|\leqq \alpha^{1/p}e^{\alpha|y|}\left|\kappa(t)\right|_{p'}\left|f(t)\right|_p<Ae^{\alpha|y|}.$$

Let C be the circle $\zeta=2Re^{i\phi}$ $(0\leqq\phi\leqq 2\pi)$, and let $|z|\leqq R$. Then

$$(4.4)\qquad \begin{aligned}\frac{1}{2\pi i}\int_C\frac{g_\alpha(\zeta)}{\zeta-z}d\zeta&=\frac{\alpha}{2\pi i}\int_{-\infty}^{\infty}f(t)dt\int_C\frac{\kappa\{\alpha(t-\zeta)\}}{\zeta-z}d\zeta\\&=\alpha\int_{-\infty}^{\infty}\kappa\{\alpha(t-z)\}f(t)dt=g_\alpha(z),\end{aligned}$$

the interchange of integrations being justified by absolute convergence. Hence $g_\alpha(z)$ is an analytic function for $|z|\leqq R$. Since R is arbitrary, $g_\alpha(z)$ is an integral function, which completes the proof.

LEMMA 5. *Let k be a positive integer, let $\kappa(z)\in G_1$ and $t^k\kappa(t)\in L_1(-\infty,\infty)$. Let $f(t)$ be measurable and bounded in any finite interval and $f(t)=O(|t|^k)$ $(t\to\pm\infty)$. Then the function $g_\alpha(z)$, defined by* (4.1), *belongs to G_α.*

Obviously $\kappa(t)\in L_1(-\infty,\infty)$. We have $|f(t)|\leqq A(1+|t|^k)\leqq A'(1+|z|^k+|t-z|^k)$ for $-\infty<t<\infty$. Therefore, taking $\kappa_0(z)=z^k\kappa(z)$, we deduce that

$$(4.5)\qquad \begin{aligned}|g_\alpha(z)|&\leqq\alpha^{1-k}A'\int_{-\infty}^{\infty}|\kappa_0\{\alpha(t-z)\}|\,dt\\&\quad+A'(|z|^k+1)\int_{-\infty}^{\infty}|\kappa\{\alpha(t-z)\}|\,dt,\\|g_\alpha(z)|&\leqq A''(1+|z|^k)e^{\alpha|y|},\end{aligned}$$

using (4.2), with $p'=1$. It is easy to show that (4.4) holds which completes the proof.

LEMMA 6. *Let $1\leqq q<\infty$, let $k-1$ be a positive integer and $p=qk^{-1}$, let $\phi(x)$, $\psi(x)\in L_q(-\infty,\infty)$ and $M=\max\{|\phi(x)|_q,|\psi(x)|_q\}$. Then*

$$|\{\phi(x)\}^k-\{\psi(x)\}^k|_p\leqq k^\lambda M^{k-1}|\phi(x)-\psi(x)|_q\quad(\lambda=\max(1,(k-1)/q)).$$

Proof. Applying Hölder's inequality, we have

$$\begin{aligned}\int_{-\infty}^{\infty}|\phi^k-\psi^k|^p dx&=\int_{-\infty}^{\infty}\left(|\phi-\psi|\left|\sum_{j=0}^{k-1}\phi^j\psi^{k-1-j}\right|\right)^p dx\\&\leqq|\phi-\psi|_q^p\left|\sum_{j=0}^{k-1}\phi^j\psi^{k-1-j}\right|_{q/(k-1)}^p,\\|\phi^j\psi^{k-1-j}|_{q/(k-1)}&\leqq(|\phi|_q^{j/(k-1)}|\psi|_q^{1-j/(k-1)})^{k-1}\leqq M^{k-1},\end{aligned}$$

which gives the required result. Obviously it holds in $(0,\infty)$.

5. **Preliminary approximation theorems.**

THEOREM 1. *Let $0<p\leqq\infty$, let $f(t)$ belong to $L_p(-\infty,\infty)$ or $L_p(0,\infty)$, let $\rho<1$ or $\rho<\frac{1}{2}$, respectively. Let $f(t)$ be approximated by elements $g_n(z)$ of $G^{(\rho)}$ $(n=1, 2, \cdots)$, that is,*

$$(5.1)\qquad |f(t)-g_n(t)|_p\to 0 \quad or \quad \|f(t)-g_n(t)\|_p\to 0 \quad as \quad n\to\infty$$

for $\rho<1$ or $\rho<\frac{1}{2}$, respectively. Then $f(t)$ reduces to a constant which is zero for $0<p<\infty$.

I. *The interval* $(-\infty,\infty)$. Here the theorem is a corollary of a result due to V. G. Iyer([12]). It can also be deduced from the Lemmas 2'(a) and 2 by means of the Liouville theorem.

II. *The interval* $(0,\infty)$. Since $\rho<\frac{1}{2}$, we have $g_n(z)\in G_0^{(1/2)}$. Hence the even function $F(z)=g_n(z^2)$ belongs to G_0. Since $g_n(t)\in L_p(0,\infty)$, we have $|t|^{1/p}F_n(t)\in L_p(-\infty,\infty)$ and, therefore, $F_n(t)\in L_p(-\infty,\infty)$. We can again employ the Lemmas 2'(a) and 2, which proves the theorem.

THEOREM 2. *Let $0<p\leqq\infty$ and let $f(t)$ belong to $L_p(-\infty,\infty)$ or $L_p(0,\infty)$, respectively, let $f(t)$ be approximated by elements $g_n(z)$ of $G^{(\rho)}$ $(n=1, 2, \cdots)$, and let σ be any number greater than ρ. Then $f(t)$ can be approximated by integral functions of order σ.*

I. *The interval* $(-\infty,\infty)$. By Theorem 1, we may suppose that $\rho\geqq 1$, $\sigma>1$. We need only find a single integral function $W(z)$ of order σ such that $W(t)\in L_p(-\infty,\infty)$ $(0<p\leqq\infty)$. For (5.1) implies that $|f(t)-h_n(t)|_p\to 0$ as $n\to\infty$, where $h_n(z)=g_n(z)+n^{-1}W(z)$, and the integral function $h_n(z)$ is of order σ.

Let $0<\lambda<1$. The Mittag-Leffler function

$$(5.2)\qquad E_\lambda(z)=\sum_{n=0}^{\infty}\frac{z^n}{\Gamma(1+\lambda n)}$$

is well known to be an integral function of the type unity of the order λ^{-1}. Let $0<\epsilon<1-\lambda$. Then, uniformly in the angle $(\pi/2)(\lambda+\epsilon)\leqq\arg z\leqq 2\pi-(\pi/2)(\lambda+\epsilon)$, we have

$$(5.3)\qquad E_\lambda(z)=O(|z|^{-1}) \qquad (|z|\to\infty).$$

When k is any integer greater than ρ^{-1} then the function $W(z)=(E_{1/\sigma}(iz))^k$ has the required property.

II. *The interval* $(0,\infty)$. It will suffice to take $\sigma>1/2$. The integral function $W(z)=(E_{1/\sigma}(-z))^k$ is an integral function of order σ, and $W(t)\in L_p(0,\infty)$, which proves the theorem.

([12]) Quart. J. Math. Oxford Ser. vol. 7 (1936) pp. 294–299, Theorems 2 and 4.

6. Approximation functions of lowest order. Let $0<p\leqq\infty$ and $f(t)\in L_p(-\infty, \infty)$, and let $f(t)$ not reduce to a constant. By Theorems 1 and 2 there exists a lower bound of the ρ's, provided that $f(t)$ can be approximated at all by integral functions of finite order. This minimum is in fact attained.

THEOREM 3(a). *Let $0<p<\infty$ and $f(t)$ belong to $L_p(-\infty, \infty)$ or $L_p(0, \infty)$, respectively. Then there are functions $g_\alpha(z)\in G_\alpha$ or $\in G_\alpha^{(1/2)}$ $(0<\alpha\to\infty)$ such that*

$$|f(t)-g_\alpha(t)|_p\to 0 \quad \text{or} \quad \|f(t)-g_\alpha(t)\|_p\to 0 \quad \text{as} \quad \alpha\to\infty.$$

THEOREM 3(b). *Let $f(t)$ belong to $L_\infty(-\infty, \infty)$ or $L_\infty(0, \infty)$, respectively. A necessary and sufficient condition for the existence of functions $g_\alpha(z)\in G_\alpha$ or $\in G_\alpha^{(1/2)}$ $(0<\alpha\to\infty)$ such that*

$$|f(t)-g_\alpha(t)|_\infty\to 0 \quad \text{or} \quad \|f(t)-g_\alpha(t)\|_\infty\to 0 \quad \text{as} \quad \alpha\to\infty$$

is that $f(t)$ or $f(t^2)$, respectively, is uniformly continuous in $(-\infty, \infty)$.

The following theorems concern classes of functions $f(t)$ which need not belong to any $L_p(-\infty, \infty)$ $(0<p\leqq\infty)$.

THEOREM 3(c). *Let, for some r $(0<r\leqq 1)$, and uniformly for $-\infty<t_j<\infty$ $(j=1, 2)$, $f(t)$ satisfy the Lipschitz condition*

$$|f(t_1)-f(t_2)|\leqq A|t_1-t_2|^r. \tag{6.1}$$

Then $f(t)$ can be approximated by functions $g_\alpha(z)\in G_\alpha$ $(0<\alpha\to\infty)$, uniformly in $(-\infty, \infty)$.

THEOREM 3(d). (Bernstein's theorem)[13]. *The preceding result holds if $f(t)$ is continuous and $f(t)-f_1(t)\to 0$ as $t\to\infty$ and $f(t)-f_2(t)\to 0$ as $t\to-\infty$, where $f_1(t)$ and $f_2(t)$ are linear functions.*

7. Proofs of Theorems 3.

I. *The interval* $(-\infty, \infty)$. We start from the following result[14].

LEMMA 7. *Let $f(t)\in L_p(-\infty, \infty)$ $(1\leqq p<\infty)$, or let $f(t)$ belong to $L_\infty(-\infty, \infty)$ and be uniformly continuous in $(-\infty, \infty)$; let $\kappa(t)\in L_1(-\infty, \infty)$ and $\int_{-\infty}^{\infty}\kappa(t)dt=1$, and let $g_\alpha(x)$ be defined by (4.1), replacing z by x. Then, as $\alpha\to\infty$, $|f(t)-g_\alpha(t)|_p\to 0$ $(1\leqq p\leqq\infty)$.*

Let $\kappa(z)$ satisfy the conditions both of Lemma 4 and of Lemma 7. Take, for instance, the Féjer kernel $\kappa(z)=(\pi z^2)^{-1}(1-\cos z)$. Then Theorem 3(a), for $1\leqq p<\infty$, and the sufficiency of the conditions in Theorem 3(b) is an immediate consequence of the lemma. The necessity of the condition in 3(b) is readily deduced from the Bernstein[15] theorem:

[13] Loc. cit., 106, footnote. It can be deduced from Theorems 3(b) and (c).

[14] E. Hille, Trans. Amer. Math. Soc. vol. 39 (1936) pp. 131–153 §1.

[15] Loc. cit., 102, Corollary 7.

LEMMA 8. *Let $g(z)\in G_\alpha$ and $|g(t)|\leqq A$ in $(-\infty, \infty)$. Then $|g'(t)|\leqq \alpha A$ in $(-\infty, \infty)$, therefore $g(t)$ is uniformly continuous in $(-\infty, \infty)$.*

Let now $0<p<1$ and $f(t)\in L_p(-\infty, \infty)$. Let k be an integer not smaller than p^{-1} and $kp=q$. Defining $\{f(t)\}^{1/k}$ in a suitable way, we have $\{f(t)\}^{1/k}\in L_q(-\infty, \infty)$. Let $\kappa(z)$ be defined as above, and let

$$(7.1)\qquad g_\alpha(z)=\left(\frac{\alpha}{k}\int_{-\infty}^{\infty}\kappa\left(\frac{\alpha t}{k}-\frac{\alpha z}{k}\right)\{f(t)\}^{1/k}dt\right)^k=\{\gamma_\beta(z)\}^k,$$

where $\beta=\alpha k^{-1}$. Then $|f(t)-g_\alpha(t)|_p\to 0$ as $\alpha\to\infty$. Obviously $\gamma_\beta(z)\in G_\beta$, and $\epsilon_\beta=|\{f(t)\}^{1/k}-\gamma_\beta(t)|_q\to 0$ as $\beta\to\infty$. By Lemma 6, we have

$$|f(t)-\{\gamma_\beta(t)\}^k|_p\leqq A\,|\{f(t)\}^{1/k}-\gamma_\beta(t)|_q\to 0 \quad \text{as} \quad \beta\to\infty,$$

since

$$\max\left(|f^{1/k}|_q,\ |\gamma_\beta(t)|_q\right)\leqq |f^{1/k}|_q+\epsilon_\beta=|f(t)|_p^{1/k}+\epsilon_\beta<A'.$$

Hence $|f(t)-g_\alpha(t)|_p\to 0$ as $\alpha\to\infty$, and obviously $g_\alpha(z)\in G_\alpha$.

To prove Theorem 3(c), we take, for instance,

$$(7.2)\qquad \kappa(z)=az^{-4}\sin^4\frac{z}{4},\qquad a^{-1}=\int_{-\infty}^{\infty}\frac{dt}{t^4}\sin^4\frac{t}{4}=\frac{\pi}{96}.$$

Since $\kappa(z)$ and, by (6.1), $f(t)$, with $k=1$, satisfy the hypotheses of Lemma 5, the function $g_\alpha(z)$, defined by (4.1), belongs to G_α. We have

$$f(x)-g_\alpha(x)=\alpha\int_{-\infty}^{\infty}\kappa(\alpha t)\{f(x)-f(t+x)\}dt.$$

Hence, by (6.1),

$$|f(x)-g_\alpha(x)|\leqq A\alpha\int_{-\infty}^{\infty}|\kappa(\alpha t)||t|^r dt=A\alpha^{-r}\int_{-\infty}^{\infty}|\kappa(t)t^r|\,dt,$$

which tends to zero as $\alpha\to\infty$, uniformly for $-\infty<x<\infty$. This completes the proof of the assertions concerning the interval $(-\infty, \infty)$.

As an application of Theorem 3(c) we obtain the following result:

Let $0<r<1$, and let z^r be uniquely defined by $-\pi/2\leqq\arg z\leqq 3\pi/2$. Then t^r can be approximated by functions $g_\alpha(z)\in G_\alpha$ $(0<\alpha\to\infty)$, uniformly in $(-\infty, \infty)$.

The result follows immediately from the inequality $|t_1^r-t_2^r|\leqq 2^{1-r}|t_1-t_2|^r$. It holds for $0<r<\infty$, and the corresponding result holds for the interval $(0, \infty)$, with $g_\alpha(z)\in G_\alpha^{(1/2)}$; this will be shown in the paper indicated above. We remark that, for $r\neq 1, 2, 3, \cdots$, t^r cannot be approximated by rational functions in $(-\infty, \infty)$ or $(0, \infty)$.

II. *The interval* $(0, \infty)$. This case is, for $p=\infty$, readily deduced from the corresponding result for the interval $(-\infty, \infty)$. We have to take into

consideration that $F(t)=f(t^2)$ is an even function and, therefore,

$$\gamma_\alpha(z)=\frac{1}{\pi\alpha}\int_{-\infty}^{\infty}\frac{1-\cos\alpha(t-z)}{(t-z)^2}F(t)dt \tag{7.3}$$

is an even function as well. Let now $p=1$ and again $F(t)=f(t^2)$. Then $tF(t)\in L_1(-\infty,\infty)$. Replacing $F(t)$ by $tF(t)$ in (7.1), we see that $\gamma_\alpha(z)$ is an odd function, and that $|tF(t)-\gamma_\alpha(t)|_1\to 0$ as $\alpha\to\infty$. Taking $g_\alpha(z)=z^{-1/2}\gamma_\alpha(z^{1/2})$, we have $g_\alpha(z)\in G_\alpha^{(1/2)}$ and $\|f(t)-g_\alpha(t)\|_1\to 0$ as $\alpha\to\infty$.

For the further investigation, we need

LEMMA 9([16]). *The sequence* $\{\pi^{-1/2}(i-t)^n(i+t)^{-n-1}\}$ $(n=0,\pm1,\pm2,\cdots)$ *is a complete orthonormal system with respect to* $L_p(-\infty,\infty)$ *for* $1<p<\infty$.

Consequently([16]), for $1<p<\infty$, any element $f(t)\in L_p(0,\infty)$, can certainly be approximated, in the mean of order p, by finite linear forms of the $(i-t)^n(i+t)^{-n-1}$. Therefore we need only show, that, given $\epsilon>0$ and any integer n, there is an element $g_\alpha(z)\in G_\alpha^{(1/2)}$ such that $\|(i-t)^n(i+t)^{-n-1}-g_\alpha(t)\|_p<\epsilon$, or that there exists an even function $\gamma_\alpha(z)\in G_\alpha$ such that

$$\int_{-\infty}^{\infty}\left|\frac{(i-t^2)^n}{(i+t^2)^{n+1}}-\gamma_\alpha(t)\right|^p|t|\,dt<\epsilon^p. \tag{7.4}$$

Let $\kappa(z)$ be defined by (7.2), and let

$$g_\alpha(z)=\alpha\int_{-\infty}^{\infty}\kappa\{\alpha(t-z)\}\frac{(i-t^2)^n}{(i+t^2)^{n+1}}dt.$$

Then, for $\alpha\geqq A=A(\epsilon)$, we have

$$\left|\frac{(i-t^2)^n}{(i+t^2)^{n+1}}-g_\alpha(t)\right|_p<\epsilon. \tag{7.5}$$

The function $g_\alpha(z)$ is even, and we have

$$\begin{aligned}xg_\alpha(x)&=-\int_{-\infty}^{\infty}\kappa_0\{\alpha(t-x)\}\frac{(i-t^2)^n dt}{(i+t^2)^{n+1}}+\alpha\int_{-\infty}^{\infty}\kappa\{\alpha(t-x)\}\frac{t(i-t^2)^n}{(i+t^2)^{n+1}}dt\\&=\phi_\alpha(x)+\psi_\alpha(x),\end{aligned}$$

where $\kappa_0(z)=z\kappa(z)$. Obviously $|t(i-t^2)^n(i+t^2)^{-n-1}-\psi_\alpha(t)|_p\to 0$ as $\alpha\to\infty$, while, by a well known convexity theorem([17]),

$$|\phi_\alpha(x)|_p\leqq\alpha^{-1}|\kappa_0(t)|_1|(i+x^2)^{-1}|_p.$$

([16]) H. Kober, a forthcoming paper in Quart. J. Math. Oxford Ser. For the case $p=2$, see also E. Hille, Compositio Math. vol. 6 (1938) p. 99.

([17]) E.g., *Inequalities*, G. H. Hardy, J. E. Littlewood and G. Pólya, Cambridge, 1934, Theorem 204.

Hence $|\phi_\alpha(x)|_p \to 0$ as $\alpha \to \infty$; consequently, for some $\alpha > A$, we have

$$|(i-t^2)^n t(i+t^2)^{-n-1} - tg_\alpha(t)|_p < \epsilon. \tag{7.6}$$

By Hölder's inequality, from (7.5) and (7.6) we deduce (7.4). For, taking $\chi(t) = (i-t^2)^n(i+t^2)^{-n-1} - g_\alpha(t)$, we have

$$\int_{-\infty}^{\infty} |\chi|^p |t|\, dt = \int_{-\infty}^{\infty} |t\chi|\, |\chi|^{p-1} dt \leqq |t\chi|_p |\chi|_p^{p-1} < \epsilon^p.$$

Thus we have proved the theorem for $1 \leqq p < \infty$. The case $0 < p < 1$ is treated as the corresponding case in $(-\infty, \infty)$.

8. **Uniqueness.** By the argument used in Theorem 2, it is evident that the sets of approximating functions, occurring in Theorems 3(a)–(c), are not at all unique. We can construct approximating functions of order $\rho = 1$ or $\rho = \frac{1}{2}$, respectively, with types tending to infinity, even when $f(t)$ reduces to a null-function. Certainly for $f(t) \in L_p(-\infty, \infty)$ or $\in L_p(0, \infty)$, however, the types tend necessarily to infinity provided that $f(t)$ is not equivalent to a constant or to an integral function of finite type of the order $\rho = 1$ or $\rho = \frac{1}{2}$, respectively. We can prove

THEOREM 4. *Let $0 < p \leqq \infty$ and $f(t) \in L_p(-\infty, \infty)$ or $\in L_p(0, \infty)$, respectively, let $f(t)$ be approximated by functions $g_\alpha(z) \in G_\alpha$ or $\in G_\alpha^{(1/2)}$; let β_α be the actual*[18] *type of $g_\alpha(z)$, and let $c = \liminf \beta_\alpha$ $(\alpha \to \infty)$ be finite. Then there is an element $F(z) \in G_c$ or $\in G_c^{(1/2)}$, respectively, such that $f(t) \equiv F(t)$ in $(-\infty, \infty)$ or in $(0, \infty)$.*

By the hypothesis, there is a sequence $\{\alpha_n\}$ $(n = 1, 2, \cdots)$ such that $g_{\alpha_n} \to f$, $n \to \infty$, and that the sequence $\{\beta_n\}$ of the actual types of the $g_{\alpha_n}(z)$ tends to c. Given $\epsilon > 0$, we have $\beta_n \leqq c + \epsilon$ for $n \geqq A = A(\epsilon)$, $g_{\alpha_n}(z) \in G_{c+\epsilon}$ or $G_{c+\epsilon}^{(1/2)}$, respectively, and

$$|g_{\alpha_m}(t) - g_{\alpha_n}(t)|_p \to 0 \quad \text{or} \quad \|g_{\alpha_m}(t) - g_{\alpha_n}(t)\|_p \to 0$$

as $m > n \to \infty$. By Lemma 3, the $g_{\alpha_n}(z)$ converge to a function $F(z)$ belonging to $G_{c+\epsilon}$ or $G_{c+\epsilon}^{(1/2)}$, respectively. We can take $\epsilon \to 0$. Evidently $F(t) \equiv f(t)$ which proves the theorem.

9. **The problem of best approximation.**

THEOREM 5. *Let $0 < p \leqq \infty$ and $f(t) \in L_p(-\infty, \infty)$ or $\in L_p(0, \infty)$, respectively, and let α be fixed. There exists a function $g(z) \in G_\alpha$ or $\in G_\alpha^{(1/2)}$ of best approximation to $f(t)$, satisfying the inequality $|f(t) - g(t)|_p \leqq |f(t) - h(t)|_p$ or $\|f(t) - g(t)\|_p \leqq \|f(t) - h(t)\|_p$ for any $h(z) \in G_\alpha$ or $\in G_\alpha^{(1/2)}$, respectively. For $1 < p < \infty$, $g_\alpha(z)$ is certainly unique.*

Let $f(t) \in L_p(-\infty, \infty)$, and let $h_0(z)$ be an arbitrary element of $G_{\alpha,p}$ (§3).

[18] $\beta_\alpha = \limsup_{r \to \infty} \max_{|z|=r} (\log |g_\alpha(z)|/r)$.

Denoting by E the set of elements $h(z)$ of G_α such that $|f(t)-h(t)|_p \leqq A_0 = |f(t)-h_0(t)|_p$, by the argument of §3 we deduce that, uniformly for $h(z)\in E$, $|h(z)| < A_1 e^{\alpha|y|}$. By a well known argument(19) we can now deduce the required result, using the Lemmas 3, 3′ and, in the proof of the last assertion of the theorem, Minkowski's inequality. The interval $(0, \infty)$ can be dealt with in a similar way.

THEOREM 5′. *Let $f(t)\in L_p(-\infty, \infty)$. Then the sequence of best approximation is $\{g_\alpha(z)\} = \{D_\alpha f\}$ $(0<\alpha\to\infty)$, where*

$$D_\alpha f = D_\alpha[f; z] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin \alpha(t-z)}{t-z} f(t)dt$$

is the Dirichlet singular integral.

By Lemma 4, we have $D_\alpha f\in G_\alpha$. Let $F(u)$ be the Fourier transform of $f(t)$. It is well known(20) that

$$D_\alpha[f; x] = (2\pi)^{-1/2}\int_{-\alpha}^{\alpha} F(u)e^{ixu}du.$$

Hence $\Phi(u)$, the Fourier transform of $f(x)-D_\alpha f$, vanishes for $|u|<\alpha$. By the Paley-Wiener(21) theorem, $\Psi(u)$, the Fourier transform of $D_\alpha f-h(x)$, vanishes for $|u|>\alpha$ whenever $h(z)\in G_{\alpha,2}$. Since the Fourier transform is isometric, we have

$$\int_{-\infty}^{\infty}(f(x)-D_\alpha f)\overline{(D_\alpha f-h(x))}dx = \int_{-\infty}^{\infty}\Phi(u)\overline{\Psi}(u)du = 0.$$

Consequently

$$\int_{-\infty}^{\infty}|f(x)-h(x)|^2dx = \int_{-\infty}^{\infty}|f(x)-D_\alpha f|^2dx + \int_{-\infty}^{\infty}|D_\alpha f-h(x)|^2dx.$$

The right side is greater than $|f(x)-D_\alpha f|_2^2$ when $h(x)$ is different from $D_\alpha f$, which proves the theorem. It can be shown that there exists a constant $a\geqq 2$ such that, for $a<p<\infty$ and for $1<p<a(a-1)^{-1}$, $D_\alpha f$ is not a function of best approximation of $f(t)$.

10. **A generalization of problem I.** Let C be a point set possessing the following properties:

(i) There exists an integral function $w=\Phi(z)$ which maps C on the real axis of the z-plane in a one-to-one correspondence.

(ii) The inverse function $z=\Psi(w)$ is continuous in $(-\infty, \infty)$.

(19) J. L. Walsh, loc. cit., chapter 12.

(20) E.g., E. Hille, loc. cit. footnote 14, §5.

(21) *Fourier transforms in the complex domain*, Amer. Math. Soc. Colloquium Publications, vol. 19, 1934, pp. 12–13.

Again let C' be a point set satisfying (i) and, instead of (ii), the weaker condition

(ii)' $\Psi'(w)$ exists and is finite for almost all w in $(-\infty, \infty)$. From Theorems 3(a) and (b) we can deduce

THEOREM 6. *Let $f(z)$ be bounded and uniformly continuous on C, or let, for $p=1, 1/2, 1/3, \cdots$, $f(z)\in L_p(C')$, that is to say, let $f(z)$ be measurable on C' and the integral*

$$\int_{C'} |f(z)|^p |dz| = \int_{-\infty}^{\infty} |f(\Psi(w))|^p |\Psi'(w)| |dw| \quad (p^{-1} \text{ a positive integer})$$

be finite. Then there are integral functions $h_\alpha(z)$ such that, as $\alpha\to\infty$,

$$\text{ess. u.b.}_{z \text{ on } C} |f(z) - h_\alpha(z)| \to 0 \tag{10.1}$$

or

$$\int_{C'} |f(z) - h_\alpha(z)|^p |dz| \to 0, \tag{10.2}$$

respectively, and $|h_\alpha(z)| < A_\alpha e^{\alpha|\Phi(z)|}$ or $< A_\alpha |\Phi'(z)|^{1/p} e^{\alpha|\Phi(z)|}$.

Making use of Theorem 3(c), we can, in the general case $0<p<\infty$, construct integral functions $h_\alpha(z)$ satisfying (10.2), certainly if C' satisfies (i) and the following conditions:

(ii)' $\Psi'(w)$ is bounded in $(-\infty, \infty)$.

(iii) For some r $(0<r\leqq 1)$ and uniformly in $(-\infty, \infty)$,

$$|(\Psi'(w_1))^{-1/p} - (\Psi'(w_2))^{-1/p}| \leqq A |w_1 - w_2|^r.$$

For instance, taking $w=\Phi(z)=z^2-2ia$, $a>0$, and taking C' as the upper branch of the hyperbola $xy=a$, it will be seen that, for $\frac{1}{2}\leqq p<\infty$, C' has the required properties, with $r=1$, $A=(8a)^{1/2}(4ap)^{-1}$.

THE UNIVERSITY,
EDGBASTON, BIRMINGHAM, ENGLAND.

THE TRANSFORMATION OF DOUBLE INTEGRALS

BY

R. G. HELSEL AND T. RADÓ

CHAPTER I. INTRODUCTION

1.1. Let there be given a continuous transformation T of the form

$$T: \quad x = x(u, v), \qquad y = y(u, v),$$

where $x(u, v)$, $y(u, v)$ are defined and continuous on the unit square

$$S_0: \quad 0 \leqq u \leqq 1, \qquad 0 \leqq v \leqq 1.$$

Designate the continuous image of the boundary of S_0 under T by C. We shall be concerned with the validity of the *topological transformation formula* (apparently first studied by Schauder [1](1))

$$\iint_{S_0} F(x(u, v), y(u, v))J(u, v; T)dudv = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x, y)\mu(x, y; S_0, T)dxdy, \tag{1}$$

where $F(x, y)$ is any measurable function in the xy-plane; $J(u, v; T)$ is the Jacobian, $x_u(u, v)y_v(u, v) - x_v(u, v)y_u(u, v)$; and $\mu(x, y; S_0, T)$ is equal to the topological index of the point (x, y) with respect to C if (x, y) is not on C; otherwise, $\mu(x, y; S_0, T) = 0$. The topological index of a point (x, y) with respect to the continuous curve C is defined as follows:

DEFINITION. *As (u, v) describes the boundary of S_0 once in the counterclockwise sense, $(x(u, v), y(u, v))$ describes in the xy-plane the directed closed continuous curve C, and, if (x, y) is a point not on C, the change of the continuously varying argument of the vector from (x, y) to $(x(u, v), y(u, v))$ is of the form $2k\pi$, where k is an integer (positive, negative, or zero). The integer k is called the topological index of the point (x, y) with respect to C.*

If $F(x, y) \equiv 1$, formula (1) becomes the *area formula*

$$\iint_{S_0} J(u, v; T)dudv = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mu(x, y; S_0, T)dxdy. \tag{2}$$

If the transformation T is biunique, $|\mu(x, y; S_0, T)|$ assumes only two values, 0 or 1, and formula (1) reduces to the *ordinary transformation formula*

Presented to the Society, April 24, 1943; received by the editors January 6, 1943.

(1) Numbers in brackets refer to bibliography at end of the paper.

$$\iint_{S_0} F(x(u, v), y(u, v)) \mid J(u, v; T) \mid dudv = \iint_{T(S_0^0)} F(x, y)dxdy,$$

where $T(S_0^0)$ is the image under T of S_0^0, the set of interior points of S_0.

1.2. Radó and Reichelderfer have established formula (1), under the assumption that the integral on the left exists and the locus of C is a set of planar measure zero, for a wide class of continuous transformations which they call K_3. (Many other results hold as well in class K_3. See Radó and Reichelderfer [1].) The purpose of the present paper is to show that this result implies all of the results on the transformation of double integrals in the literature of which we are aware (see McShane [1], Morrey [1], Rademacher [1], Radó [1, 2, 3], Schauder [1], Young [1, 2, 3]). It is not difficult to verify, by means of the two theorems of Radó and Reichelderfer quoted in 1.6 and 1.7, that the transformations considered by McShane [1], Morrey [1], Rademacher [1], Radó [1, 2, 3], and Schauder [1] belong to class K_3 and that the formulas established by these authors follow from formula (1). On the other hand, it is not apparent that Young's results are implied by those of Radó and Reichelderfer. We shall devote our attention, then, to the work of Young [1, 2, 3], ultimately showing that all of his results can be accounted for in terms of the results of Radó and Reichelderfer. (It is interesting to note that our methods will place many of the transformations considered by Young in Morrey's class L. Morrey [1] has established formula (2) for transformations in class L.)

1.3. The two transformation formulas which Young develops are the same on the left as formulas (1) and (2); however, the right sides of Young's formulas are not Lebesgue integrals, but the limits of Lebesgue integrals. We shall now give a brief description of these limits. Let σ be the generic notation for a subdivision of the boundary of S_0 by points $P_1, \cdots, P_m$, numbered consecutively in the positive sense around S_0. Let Π_σ be the directed closed polygon "inscribed" in C which is formed by straight segments connecting in order the images of $P_1, \cdots, P_m, P_1$ under T. Define $\mu_\sigma(x, y)$ to be equal to the topological index of the point (x, y) with respect to Π_σ if (x, y) is not on Π_σ; otherwise, set $\mu_\sigma(x, y) = 0$. Then if $F(x, y)$ is any measurable function in the xy-plane, Young defines the *integral of $F(x, y)$ over the area of the polygon* Π_σ as follows:

$$(Y) \iint_{\Pi_\sigma} F(x, y)dxdy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(x, y)\mu_\sigma(x, y)dxdy,$$

providing the Lebesgue integral on the right exists. Young then defines the *integral of $F(x, y)$ over the area of the curve C* in the following way:

$$(Y) \iint_C F(x, y)dxdy = \lim_{\|\sigma\| \to 0} (Y) \iint_{\Pi_\sigma} F(x, y)dxdy,$$

providing the limit on the right exists and is independent of the manner in which $\|\sigma\| \to 0$. ($\|\sigma\|$ is the norm of the subdivision σ.)

1.4. For certain classes of continuous transformations

$$T: \quad x = x(u, v), \qquad y = y(u, v), \qquad (u, v) \in S_0,$$

Young has established the *area formula*

$$\iint_{S_0} J(u, v; T)dudv = (Y)\iint_C dxdy, \tag{3}$$

where C is the image of the boundary of S_0 under T and $(Y)\iint_C dxdy$ is assumed to exist. In a few instances, he also asserts that the *general transformation formula*

$$\iint_{S_0} F(x(u, v), y(u, v))J(u, v; T)dudv = (Y)\iint_C F(x, y)dxdy \tag{4}$$

holds for any measurable function $F(x, y)$ in the xy-plane as soon as the integral on the right exists. Throughout his work, Young assumes that the coordinate functions $x(u, v)$ and $y(u, v)$ of the transformation T are absolutely continuous on every horizontal and on every vertical in S_0; hence, the continuous curve C is rectifiable and its locus is a set of planar measure zero.

1.5. A comparison of formulas (1) and (4) leads to this engaging question: When does

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x, y)\mu_\sigma(x, y)dxdy \xrightarrow[\|\sigma\|\to 0]{} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x, y)\mu(x, y; S_0, T)dxdy?$$

Young neither raises nor answers this question. Instead, he always assumes the existence of the limit on the left and defines it to be the new integral, $(Y)\iint_C F(x, y)dxdy$. It is clear that an improvement is achieved if in formula (4) we can replace $(Y)\iint_C F(x, y)dxdy$ by $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x, y)\mu(x, y; S_0, T)dxdy$. (Schauder [1] has already enunciated the desideratum of replacing $(Y)\iint_C F(x, y)dxdy$ by something more tangible.) In fact, in every case considered by Young but one, we shall accomplish this replacement by showing that the transformation belongs to class K_3 and, hence, formula (1) holds as soon as the integral on the left exists (cf. 1.2, 1.4). In the one exceptional case just mentioned, it would seem that Young did not establish anything (cf. 3.10). (The preceding statements should not be construed to mean that we have succeeded in identifying $(Y)\iint_C F(x, y)dxdy$ and $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x, y)\mu(x, y; S_0, T)dxdy$ in all the cases of Young. Actually, it appears to us that a general study of the relation of these integrals would lead to questions of great difficulty and interest.)

1.6. To place Young's transformations in class K_3, we shall utilize the two following theorems of Radó and Reichelderfer [1].

THEOREM. *If the transformation T is continuous on the domain $\mathcal{D}$ of the uv-plane; the partial derivatives $x_u(u, v)$, $x_v(u, v)$, $y_u(u, v)$, $y_v(u, v)$ exist and are continuous on $\mathcal{D}$; and $J(u, v; T)$ is summable on $\mathcal{D}$, then $T \in K_3(\mathcal{D})$.*

1.7. CLOSURE THEOREM. *Let there be given bounded domains $\mathcal{D}$ and $\mathcal{D}_n$ in the uv-plane and continuous transformations*

$$T: \quad x = x(u, v), \qquad y = y(u, v), \qquad (u, v) \in \mathcal{D},$$

$$T_n: \quad x = x_n(u, v), \qquad y = y_n(u, v), \qquad (u, v) \in \mathcal{D}_n,$$

with the following properties: (i) $\mathcal{D}_n \subset \mathcal{D}$ *and, for every closed set* $F \subset \mathcal{D}$, *there exists an* $n_0 = n_0(F)$ *such that* $F \subset \mathcal{D}_n$ *for all* $n > n_0$; (ii) $J(u, v; T)$ *exists a.e. (almost everywhere) and is summable on* $\mathcal{D}$; (iii) $T_n \in K_3(\mathcal{D}_n)$ *for* $n = 1, 2, \cdots$; (iv) *for every closed set* $F \subset \mathcal{D}$, $x_n(u, v) \to_n$ *uniformly to* $x(u, v)$, $y_n(u, v) \to_n$ *uniformly to* $y(u, v)$, *on* F *and*

$$\lim_n \iint_F | J(u, v; T_n) - J(u, v; T) | \, du dv = 0.$$

Then $T \in K_3(\mathcal{D})$.

(The Closure Theorem is still valid if the condition

$$\lim_n \iint_F | J(u, v; T_n) - J(u, v; T) | \, du dv = 0$$

is replaced by the weaker condition

$$\lim_n \iint_F | J(u, v; T_n) | \, du dv = \iint_F | J(u, v; T) | \, du dv.)$$

1.8. The method employed in Chapter III to place each transformation considered by Young in class K_3 is this: We approximate to the given transformation by a sequence of transformations which belong to class K_3 (to begin, we use transformations which belong to class K_3 because of 1.6) and satisfy the conditions of the Closure Theorem, thus placing the given transformation in class K_3. The feature of this paper is the utilization of integral means to obtain the approximating transformations.

The second chapter of this paper is a systematic presentation of general theorems in analysis and properties of integral means which we apply in Chapter III to prove our results. Since integral means give rise to non-denumerable sequences of approximating functions, many of the theorems in Chapter II are concerned with such sequences; however, we work only with denumerable sequences in Chapter III. The intrinsic interest of our results on non-denumerable sequences of integral means justifies their inclusion.

1.9. The cases which we treat in Chapter III are numerous and contain many assumptions so we shall not list them here, but we shall summarize

the essential differences between our results and those of Young.

1. By placing Young's transformations in class K_3, we do more than establish formula (1) of 1.1 for them, since many other results hold as well for transformations in class K_3 (cf. 1.2).

2. In most of his cases, Young did not establish formula (4) of 1.4, but only the area formula (3). On the other hand, we shall establish the general topological transformation formula (1) for all of Young's transformations. Young established formula (3) for the transformations considered in 3.6, 3.7, 3.9, 3.12, and 3.14; formula (4) for the transformations of 3.8 and 3.14.1. The transformation of 3.14.1 is the most general case for which Young claims to have established formula (4).

3. Throughout Chapter III we assume there is given a continuous transformation $T: x=x(u, v), y=y(u, v), (u, v)\in S_0$, which satisfies the standard hypothesis H_0: $x(u, v)$ and $y(u, v)$ are absolutely continuous in the Tonelli sense (cf. 2.17) on S_0 and the image of the boundary of S_0 under T is a set of planar measure zero. The additional restrictions on T which we use in Chapter III do not differ from restrictions used by Young; however, Young always makes the standard assumption that $x(u, v)$ and $y(u, v)$ are of bounded variation in the Tonelli sense (cf. 2.17) on S_0 and are absolutely continuous on every horizontal and on every vertical in S_0. Clearly our hypothesis H_0 is less restrictive than Young's standard assumption (cf. 1.4).

It must be remembered that in all cases our theorems are not identical with those of Young (cf. 1.3, 1.5). The references to Young's work which appear in 3.6, 3.7, 3.8, 3.9, 3.12, 3.14, and 3.14.1 direct the reader to the theorems of Young which correspond to our theorems of those sections.

1.10. Young's theorems on the transformation of double integrals are not final in character; hence, we have attempted to find a general theorem which would at once account for all of his results. The following proposition, which we have not been able to establish or deny, seems a likely generalization (cf. 3.15):

If T satisfies H_0 (cf. 1.9) and $J(u, v; T)$ is summable on S_0, then $T\in K_3(S_0^0)$ and formula (1) *holds if the integral on the left exists.*

Chapter II. Auxiliary theorems and integral means

2.1. Most of the results listed in this chapter are known; however, a few of them are new—for example, see 2.12, 2.21, 2.21.1. Proofs are included whenever convenient references are not available.

2.2. The following concept will be used extensively in the present paper:

Definition. *A family $\mathfrak{F}$ of measurable functions defined on S_0 is said to have the Vitali property—briefly, property $(\mathcal{V})$—on a measurable subset E of S_0^0 if the following condition is satisfied. To every $\epsilon>0$ there corresponds a $\delta=\delta(\epsilon)>0$ such that*

$$\left| \iint_e f(u, v)dudv \right| < \epsilon$$

for every function $f(u, v)$ of $\mathcal{F}$ and every measurable set $e \subset E$ with $|e| < \delta$. (If e is any set, $|e|$ denotes the exterior measure of e.)

2.2.1. COROLLARY. *If a family $\mathcal{F}$ of functions $f(u, v)$ has property $(\mathcal{U})$ on a measurable subset E of S_0, then the family $\mathcal{F}^*$ of functions $|f(u, v)|$ also has, and conversely.*

2.2.2. COROLLARY. *If $\mathcal{F}$ is a family of functions possessing property $(\mathcal{U})$ on a measurable subset E of S_0 and $\mathcal{G}$ is a family of measurable functions which are uniformly bounded on E, then the family of functions $f(u, v) \cdot g(u, v)$, $f(u, v) \in \mathcal{F}$ and $g(u, v) \in \mathcal{G}$, has property $(\mathcal{U})$ on E.*

2.2.3. COROLLARY. *If a family $\mathcal{F}$ of functions has property $(\mathcal{U})$ on a measurable subset E of S_0, then $\iint_E f(u, v)dudv$ exists for every function $f(u, v)$ of $\mathcal{F}$ and these integrals are uniformly bounded.*

2.3. THEOREM. *If on a measurable subset E of S_0, $f_n(u, v) \geqq 0$ and $f_n(u, v) \to_n f(u, v)$ a.e., a necessary and sufficient condition for $f(u, v)$ to be summable on E and*

$$\iint_E f_n(u, v)dudv \underset{n}{\to} \iint_E f(u, v)dudv$$

is that $\{f_n(u, v)\}$ have property $(\mathcal{U})$ on E.

Proof. See de la Vallée Poussin [1, p. 477]. This theorem may be extended to a non-denumerable sequence of continuous functions $f_\alpha(u, v)$, $0 < \alpha < 1$, such that $f_\alpha(u, v) \to_{\alpha \to 0} f(u, v)$ a.e. on E.

2.3.1. COROLLARY. *If on a measurable subset E of S_0, $f_n(u, v) \to_n f(u, v)$ a.e., a necessary and sufficient condition for $f(u, v)$ to be summable on E and*

$$\iint_E |f_n(u, v) - f(u, v)| dudv \underset{n}{\to} 0$$

is that $\{f_n(u, v)\}$ have property $(\mathcal{U})$ on E.

2.4. Many of the properties of integral means which we shall utilize depend upon the next two theorems concerning partial differentiation of an indefinite integral.

THEOREM. *If $f(u, v)$ is summable on S_0 and*

$$F(u, v) = \int_0^u \int_0^v f(\xi, \eta)d\xi d\eta, \qquad (u, v) \in S_0,$$

then, for every v, $F_u(u, v) = \int_0^u f(u, \eta)d\eta$, provided u does not belong to a set of measure zero which is independent of v.

Proof. See Fubini. [1]. (Similarly, for every u, $F_v(u, v) = \int_0^v f(\xi, v)d\xi$, provided v does not belong to a set of measure zero which is independent of u. A majority of the theorems of this chapter are stated in an unsymmetric form as regards u and v. In each such instance, a second theorem may be obtained by interchanging the rôles of u and v.)

2.5. THEOREM. *If $f(u, v)$ is summable on S_0 and*

$$g(u, v) = \int_0^u f(\xi, v)d\xi, \qquad (u, v) \in S_0,$$

(for a.e. v, $g(u, v)$ exists for every u) then $g_u(u, v) = f(u, v)$ a.e. on S_0.

Proof. See Helsel and Young [1].

2.6. We shall have occasion to employ the following types of integral means.

DEFINITION. *If $f(u, v)$ is summable on S_0 and $0 < h < 1/2$ is fixed, then*

$$f_h^h(u, v) = \frac{1}{4h^2}\int_{-h}^{h}\int_{-h}^{h} f(u + \alpha, v + \beta)d\alpha d\beta = \frac{1}{4h^2}\int_{u-h}^{u+h}\int_{v-h}^{v+h} f(\xi, \eta)d\xi d\eta,$$

defined on the square S_h: $h \leqq u \leqq 1-h$, $h \leqq v \leqq 1-h$, is called the h-h-integral mean of $f(u, v)$.

DEFINITION. *If $f(u, v)$ is summable on S_0 and $0 < h < 1/2$ is fixed, then*

$$f_h(u, v) = \frac{1}{2h}\int_{-h}^{h} f(u + \alpha, v)d\alpha = \frac{1}{2h}\int_{u-h}^{u+h} f(\xi, v)d\xi,$$

defined for a.e. v on the rectangle R_{h0}: $h \leqq u \leqq 1-h$, $0 \leqq v \leqq 1$, is called the h-integral mean of $f(u, v)$.

2.7. THEOREM. *If $f(u, v)$ is continuous on S_0 and R (a closed oriented rectangle) comprised in S_0^0 is fixed, then, as $h \to 0$, $f_h^h(u, v) \to$ uniformly to $f(u, v)$ on R.*

Proof. Well known properties of integral means such as this are established in Bray [1], Morrey [1], and Radó [2]. When theorems on integral means are stated without proofs or references, the reader should consult the above papers.

2.8. THEOREM. *If $f(u, v)$ is continuous on S_0, then on S_h*

$$\frac{\partial f_h^h(u, v)}{\partial u} = \frac{1}{4h^2}\int_{v-h}^{v+h} \{f(u + h, \eta) - f(u - h, \eta)\}d\eta.$$

2.9. THEOREM. *If $f(u, v)$ is summable on S_0, then $f_h^h(u, v) \to_{h \to 0} f(u, v)$ a.e. on S_0.*

Proof. The result is immediate from Lebesgue's theorem on absolutely continuous set functions. See de la Vallée Poussin [1].

2.10. THEOREM. *If $f(u, v)$ is summable on S_0 and $R \subset S_0^0$ is fixed, then $\{f_h^h(u, v)\}$ has property $(\mathcal{U})$ on R.*

Proof. Let e be any measurable set comprised in R. Then for h small enough that $R \subset S_h$, we have

$$\left| \iint_e f_h^h(u, v)\,du\,dv \right| \leqq \frac{1}{4h^2} \int_{-h}^{h} \int_{-h}^{h} \left\{ \iint_e | f(u + \alpha, v + \beta) |\, du\,dv \right\} d\alpha\,d\beta$$

$$\leqq \frac{1}{4h^2} \int_{-h}^{h} \int_{-h}^{h} \left\{ \iint_{e_\alpha^\beta} | f(u, v) |\, du\,dv \right\} d\alpha\,d\beta$$

$$\leqq \frac{1}{4h^2} \int_{-h}^{h} \int_{-h}^{h} \Omega(|f|, S_0, |e|)\,d\alpha\,d\beta = \Omega(|f|, S_0, |e|),$$

where

$$e_\alpha^\beta = \underset{(u,v)}{E} [(u - \alpha, v - \beta) \in e],$$

$$\Omega(|f|, S_0, |e|) = \max_{E \text{ meas} \subset S_0, |E| \leq |e|} \iint_E | f(u, v) |\, du\,dv.$$

Because of the absolute continuity of $\iint_E |f(u, v)|\,du\,dv$ as a set function on S_0, $\Omega(|f|, S_0, |e|) \to_{|e| \to 0} 0$; hence, $\{f_h^h(u, v)\}$ has property $(\mathcal{U})$ on R.

2.11. THEOREM. *If $f(u, v)$ and $g(u, v)$ belong to associated Lebesgue classes L^p and L^q respectively, where $p > 0$, $q > 0$, $1/p + 1/q = 1$, and if $R \subset S_0^0$ is fixed, then $\{f_h^h(u, v) \cdot g_h^h(u, v)\}$ has property $(\mathcal{U})$ on R.*

Proof. The result follows from Hölder's inequality and reasoning similar to that employed in 2.10.

2.12. THEOREM. *If $f(u, v)$ is summable on S_0 and the closed interval (a, b) is strictly interior to $(0, 1)$, then*

$$\int_a^b | f_h^h(u, v) |\, du \xrightarrow[h \to 0]{} \int_a^b | f(u, v) |\, du,$$

provided v does not belong to a set of measure zero which is independent of a and b.

Proof. By 2.9, $f_h^h(u, v) \to_{h \to 0} f(u, v)$ a.e. on S_0; hence from the Theorem of Fubini, $f_h^h(u, v) \to_{h \to 0} f(u, v)$ a.e. on a.e. line $v = \text{constant}$ in S_0. Let E_v^1 be the exceptional set of v values. Also, by the Theorem of Fubini, $\int_0^1 |f(u, v)|\,du$

exists for v not belonging to an exceptional set E_v^2 of measure zero.

Choose two rational numbers α and β such that $0<\alpha<\beta<1$. Then define

$$g(v) = \int_{\alpha-\delta}^{\beta+\delta} |f(u, v)|\, du, \qquad v \text{ not} \in E_v^2,$$

where $\delta>0$ is rational and $0\leqq\alpha-\delta<\beta+\delta\leqq 1$. Since $g(v)$ is summable on $0\leqq v\leqq 1$, it follows from Lebesgue's theorem on absolutely continuous set functions that

$$\frac{1}{2h}\int_{-h}^{h} g(v+\eta)d\eta \xrightarrow[h\to 0]{} g(v),$$

for v not belonging to an exceptional set $E_v^3(\alpha, \beta, \delta)$ of measure zero. Put

$$E_v = E_v^1 + E_v^2 + \sum_{\alpha,\beta,\delta \text{ rat.}} E_v^3(\alpha, \beta, \delta).$$

Then $|E_v| = 0$.

For v not$\in E_v$ and for a fixed rational $\delta>h$,

$$\begin{aligned}\int_\alpha^\beta |f_h(u, v)|\, du &\leqq \frac{1}{4h^2}\int_{-h}^{h}\int_{-h}^{h}\left\{\int_\alpha^\beta |f(u+\xi, v+\eta)|\, du\right\}d\xi d\eta \\ &\leqq \frac{1}{4h^2}\int_{-h}^{h}\int_{-h}^{h}\left\{\int_{\alpha-\delta}^{\beta+\delta} |f(u, v+\eta)|\, du\right\}d\xi d\eta \\ &\leqq \frac{1}{2h}\int_{-h}^{h} g(v+\eta)d\eta.\end{aligned}$$

Therefore

$$\limsup_{h\to 0}\int_\alpha^\beta |f_h(u, v)|\, du \leqq g(v) = \int_{\alpha-\delta}^{\beta+\delta} |f(u, v)|\, du, \quad v \text{ not} \in E_v.$$

As $\delta\to 0$ through positive rational values, we obtain

$$\limsup_{h\to 0}\int_\alpha^\beta |f_h(u, v)|\, du \leqq \int_\alpha^\beta |f(u, v)|\, du, \qquad v \text{ not} \in E_v.$$

This inequality holds for any two real numbers $0<a<b<1$, since for v not $\in E_v$

$$\begin{aligned}\limsup_{h\to 0}\int_a^b |f_h(u, v)|\, du &\leqq \limsup_{h\to 0}\int_{\alpha_n}^{\beta_n} |f_h(u, v)|\, du \\ &\leqq \int_{\alpha_n}^{\beta_n} |f(u, v)|\, du \underset{n}{\to} \int_a^b |f(u, v)|\, du,\end{aligned} \tag{1}$$

where $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of rational numbers converging to a and b respectively and satisfying $0<\alpha_n<a<b<\beta_n<1$.

By the Lemma of Fatou, for any two real numbers $0<a<b<1$

$$\int_a^b |f(u,v)|\,du = \int_a^b \liminf_{h\to 0} |f_h(u,v)|\,du$$

$$\leqq \liminf_{h\to 0} \int_a^b |f_h(u,v)|\,du, \quad v \text{ not} \in E_v^1 + E_v^2. \tag{2}$$

The result follows from (1) and (2).

2.13. LEMMA. *On $a\leqq u\leqq b$, let there be given a family of functions $f_\alpha(u)$, $0<\alpha<1$, which converge a.e. to $f(u)$ as $\alpha\to 0$. If, for $0<\alpha<1$ and $a\leqq u\leqq b$, $|f_\alpha(u)|\leqq g(u)$, where $g(u)$ is summable on $a\leqq u\leqq b$, then $\int_a^b f_\alpha(u)du\to_{\alpha\to 0}\int_a^b f(u)du$.*

Proof. Deny the truth of the assertion. Then there exists a denumerable subsequence of functions $\{f_{\alpha_n}(u)\}$ such that $f_{\alpha_n}(u)\to_n f(u)$ a.e. on $a\leqq u\leqq b$ and $\int_a^b f_{\alpha_n}(u)du$ does not converge to $\int_a^b f(u)du$. This contradicts Lebesgue's theorem on termwise integration.

2.14. THEOREM. *If $f(u,v)\geqq 0$ is summable on S_0, $\lambda(u)\geqq 0$ is summable on $0\leqq u\leqq 1$, and $V(u)=\int_0^1 f(u,v)dv$ is a bounded function of u (on the set where it exists), then for a fixed rectangle R: $0<a\leqq u\leqq b<1$, $0<c\leqq v\leqq d<1$,*

$$\iint_R f_h(u,v)\lambda(u)dudv \xrightarrow[h\to 0]{} \iint_R f(u,v)\lambda(u)dudv.$$

Proof. Choose h small enough that $R\subset S_h$. By 2.12, for a.e. u, $a\leqq u\leqq b$,

$$\lambda(u)\int_c^d f_h(u,v)dv \xrightarrow[h\to 0]{} \lambda(u)\int_c^d f(u,v)dv.$$

Also

$$\lambda(u)\int_c^d f_h(u,v)dv = \frac{\lambda(u)}{4h^2}\int_{-h}^{h}\int_{-h}^{h}\left\{\int_c^d f(u+\alpha, v+\beta)dv\right\}d\alpha d\beta$$

$$\leqq \frac{\lambda(u)}{4h^2}\int_{-h}^{h}\int_{-h}^{h} V(u+\alpha)d\alpha d\beta \leqq M\cdot\lambda(u),$$

where M is a bound for $V(u)$. By the lemma of 2.13

$$\int_a^b \left\{\lambda(u)\int_c^d f_h(u,v)dv\right\}du \xrightarrow[h\to 0]{} \int_a^b \left\{\lambda(u)\int_c^d f(u,v)dv\right\}du.$$

Hence, by the Theorem of Tonelli,

$$\iint_R f_h(u,v)\lambda(u)dudv \xrightarrow[h\to 0]{} \iint_R f(u,v)\lambda(u)dudv.$$

2.15. THEOREM. *If $f(u, v)$ is continuous on S_0, then on R_{h0} (cf. 26)*

$$\frac{\partial f_h(u, v)}{\partial u} = \frac{1}{2h}\{f(u + h, v) - f(u - h, v)\}.$$

2.16. THEOREM. *If $f(u, v)$ is summable on S_0, then $f_h(u, v) \to_{h\to 0} f(u, v)$ a.e. on S_0.*

Proof. The result is immediate from 2.5.

2.17. For the purpose of Chapter III, we shall need certain theorems pertaining to functions which are *absolutely continuous in the sense of Tonelli*—briefly, A.C.T.—and to their integral means.

DEFINITION. *The function $f(u, v)$ is* A.C.T. *on S_0 if* (i) *$f(u, v)$ is continuous*; (ii) *$f(u, v)$ is of bounded variation in the sense of Tonelli*; (iii) *for a.e. $v = \eta$, $0 \leqq \eta \leqq 1$, the function $f(u, \eta)$ is absolutely continuous in u, and, for a.e. $u = \xi$, $0 \leqq \xi \leqq 1$, the function $f(\xi, v)$ is absolutely continuous in v. Denote by $V(\eta; f)$ the total variation of $f(u, \eta)$ as a function of u on $0 \leqq u \leqq 1$; denote by $V(\xi; f)$ the total variation of $f(\xi, v)$ as a function of v on $0 \leqq v \leqq 1$. Then $f(u, v)$ is of bounded variation in the sense of Tonelli if $V(\eta; f)$ and $V(\xi; f)$ are summable functions of η and ξ respectively on the interval* (0, 1).

2.18. If $f(u, v)$ is A.C.T. on S_0, the partial derivatives $f_u(u, v)$ and $f_v(u, v)$ exist a.e. and are summable on S_0; moreover, for a.e. v, $V(v; f) = \int_0^1 |f_u(u, v)|\, \partial u$ and, for a.e. u, $V(u; f) = \int_0^1 |f_v(u, v)|\, dv$ (see Morrey [1]).

2.19. THEOREM. *If $f(u, v)$ is A.C.T. on S_0, then $\partial f_h^h(u, v)/\partial u = (\partial f(u, v)/\partial u)_h^h$ on S_h.*

2.19.1. COROLLARY. *If $f(u, v)$ is A.C.T. on S_0, then $\partial f_h^h(u, v)/\partial u \to_{h\to 0} \partial f(u, v)/\partial u$ a.e. on S_0.*

Proof. The result is immediate from 2.19 and 2.9.

2.20. THEOREM. *If $f(u, v)$ is A.C.T. on S_0, then $\partial f_h(u, v)/\partial u = (\partial f(u, v)/\partial u)_h$ on R_{h0} providing v does not belong to a set of measure zero which is independent of h.*

Proof. By 2.15, $\partial f_h(u, v)/\partial u = (1/2h)\{f(u+h, v) - f(u-h, v)\}$ on R_{h0} Since $f(u, v)$ is A.C.T., it is absolutely continuous in u for a.e. v, $0 \leqq v \leqq 1$. Fix a value of v for which $f(u, v)$ is absolutely continuous in u. Then

$$\left(\frac{\partial f(u, v)}{\partial u}\right)_h = \frac{1}{2h}\int_{u-h}^{u+h} \frac{\partial f(\xi, v)}{\partial \xi}\, d\xi$$

$$= \frac{1}{2h}\{f(u + h, v) - f(u - h, v)\} = \frac{\partial f_h(u, v)}{\partial u}.$$

2.20.1. COROLLARY. *If $f(u,v)$ is A.C.T. on S_0, then $\partial f_h(u,v)/\partial u \to_{h\to 0} \partial f(u,v)/\partial u$ a.e. on S_0.*

Proof. The result is immediate from 2.20 and 2.16.

2.21. THEOREM. *If $f(u,v)$ is A.C.T. on S_0, then $\partial f_h(u,v)/\partial v = (\partial f(u,v)/\partial v)_h$ on R_{h0} providing v does not belong to a set of measure zero which is independent of h.*

Proof. For $0 \leqq v \leqq 1$,

$$\begin{aligned}\int_0^v \left(\frac{\partial f(u,\eta)}{\partial \eta}\right)_h d\eta &= \frac{1}{2h}\int_{-h}^{h}\left\{\int_0^v \frac{\partial f(u+\alpha,\eta)}{\partial \eta}\,d\eta\right\}d\alpha \\ &= \frac{1}{2h}\int_{-h}^{h}\{f(u+\alpha,v) - f(u+\alpha,0)\}\,d\alpha \\ &= f_h(u,v) - f_h(u,0).\end{aligned}$$

Therefore, on R_{h0},

$$f_h(u,v) = f_h(u,0) + \frac{1}{2h}\int_0^v \int_{u-h}^{u+h} \frac{\partial f(\xi,\eta)}{\partial \eta}\,d\xi d\eta.$$

By 2.4,

$$\frac{\partial f_h(u,v)}{\partial v} = \frac{1}{2h}\int_{u-h}^{u+h} \frac{\partial f(\xi,v)}{\partial v}\,d\xi = \left(\frac{\partial f(u,v)}{\partial v}\right)_h$$

on R_{h0} providing v does not belong to a set of measure zero which is independent of h.

2.21.1. COROLLARY. *If $f(u,v)$ is A.C.T. on S_0, then $\partial f_h(u,v)/\partial v \to_{h\to 0} \partial f(u,v)/\partial v$ a.e. on S_0.*

2.22. THEOREM. *If $f(u,v)$ is A.C.T. on S_0, then $f_h(u,v)$ is A.C.T. on R_{h0}.*

Proof. Because of the absolute continuity of the indefinite Lebesgue integral, $f_h(u,v)$ is absolutely continuous in u for every v, $0 \leqq v \leqq 1$; hence,

$$\begin{aligned}V(v; f_h) &= \int_h^{1-h} \left|\frac{\partial f_h(u,v)}{\partial u}\right| du \\ &= \frac{1}{2h}\int_h^{1-h} |f(u+h,v) - f(u-h,v)|\,du \leqq \frac{M}{h},\end{aligned}$$

where M is a bound for $|f(u,v)|$ on S_0. Thus $V(v; f_h)$ is summable on $0 \leqq v \leqq 1$.

From the proof of 2.21, on R_{h0},

$$f_h(u,v) = f_h(u,0) + \frac{1}{2h}\int_0^v \left\{\int_{u-h}^{u+h} \frac{\partial f(\xi,\eta)}{\partial \eta}\,d\xi\right\}d\eta.$$

Therefore, $f_h(u, v)$ is absolutely continuous in v for every u, $h \leqq u \leqq 1-h$, and

$$V(u; f_h) = \int_0^1 \left| \frac{\partial f_h(u, v)}{\partial v} \right| dv = \int_0^1 \left| \left(\frac{\partial f(u, v)}{\partial v} \right)_h \right| dv$$

$$\leqq \frac{1}{2h} \int_0^1 \left\{ \int_{u-h}^{u+h} | f_v(\xi, v) | \, d\xi \right\} dv$$

$$\leqq \frac{1}{2h} \iint_{S_0} | f_v(u, v) | \, du dv.$$

Thus $V(u; f_h)$ is summable on $h \leqq u \leqq 1-h$ and the proof is complete.

CHAPTER III. TRANSFORMATION FORMULAS FOR DOUBLE INTEGRALS

3.1. In this final chapter we shall take each transformation for which W. H. Young established either formula (3) or (4) of 1.4 and, by placing it in class K_3, show that formula (1) of 1.1 holds if the integral on the left exists (cf. 1.2). Throughout the chapter we make the assumption that there is given a transformation

$$T: \quad x = x(u, v), \qquad y = y(u, v), \qquad (u, v) \in S_0,$$

which satisfies hypothesis H_0 of 1.9. Additional restrictions on the transformation T will be stated in the sequel. In the last three theorems of this chapter (cf. 3.9, 3.12, 3.14), the conditions imposed on the coordinate functions $x(u, v)$ and $y(u, v)$ are not symmetric. A second theorem may be obtained in each of these cases by interchanging the rôles of $x(u, v)$ and $y(u, v)$.

3.2. We shall now show that certain transformations obtained from T by approximating to the functions $x(u, v)$, $y(u, v)$ with h-h-integral means belong to class K_3. First we introduce the following notations:

$${}_n^n f(u, v) = f_{1/n}^{1/n}(u, v), \qquad {}_n f(u, v) = f_{1/n}(u, v), \qquad {}_n S = S_{1/n}, \qquad n = 3, 4, \cdots,$$

where $f(u, v)$ is any summable function defined on S_0.

3.3. LEMMA. *If T satisfies H_0, the transformation*

$${}_n^n T_m^m: \quad x = {}_n^n x(u, v), \qquad y = {}_m^m y(u, v), \qquad n, m > 2, (u, v) \in {}_n S \cdot {}_m S,$$

belongs to class K_3 on ${}_n S^0 \cdot {}_m S^0$.

Proof. By 2.8, $\partial_n^n x(u, v)/\partial u$, $\partial_n^n x(u, v)/\partial v$, $\partial_m^m y(u, v)/\partial u$, $\partial_m^m y(u, v)/\partial v$ are continuous on ${}_n S \cdot {}_m S$; hence, ${}^n J^m = J(u, v; {}_n^n T_m^m)$ is summable there. By 1.6, ${}_n^n T_m^m \in K_3({}_n S^0 \cdot {}_m S^0)$.

3.3.1. COROLLARY. *If T satisfies H_0, the transformation*

$${}_n^n T_n^n: \quad x = {}_n^n x(u, v), \qquad y = {}_n^n y(u, v), \qquad n > 2, (u, v) \in {}_n S,$$

belongs to class K_3 on ${}_nS^0$.

3.3.2. COROLLARY. *If T satisfies H_0, the transformation ${}_n^nT_m^m$ $(1/n+1/m<1/2)$ belongs to class K_3 on $S^0_{1/n+1/m}$.*

Proof. From the definition of class $K_3(\mathcal{D})$, if $T\in K_3(\mathcal{D})$ and $\mathcal{D}^*$ is any domain within $\mathcal{D}$, then $T\in K_3(\mathcal{D}^*)$. See Radó and Reichelderfer [1].

3.4. LEMMA. *If T satisfies H_0, the transformation*

$$ {}_n^nT: \quad x = {}_n^nx(u, v), \qquad y = y(u, v), \qquad\qquad n > 2, (u, v) \in {}_nS, $$

belongs to class K_3 on ${}_nS^0$.

Proof. Fix a value of $n>2$. For each positive integer m, satisfying $1/n+1/m<1/2$, define the transformation ${}_n^nT_m^m$ on $S_{1/n+1/m}$. We shall verify that the transformations ${}_n^nT$ and ${}_n^nT_m^m$ satisfy the conditions of the Closure Theorem (cf. 1.7).

(i) Clearly $S^0_{1/n+1/m}\subset {}_nS^0$ and, for every closed set $F\subset {}_nS^0$, there exists an $m_0=m_0(F)$ such that $F\subset S^0_{1/n+1/m}$ for all $m>m_0$.

(ii) By 2.8, $\partial_n^nx(u, v)/\partial u$ and $\partial_n^nx(u, v)/\partial v$ are continuous on ${}_nS$ and, therefore, bounded. The summability of ${}^nJ=J(u, v; {}_n^nT)$ on ${}_nS^0$ follows then from the summability of $y_u(u, v)$ and $y_v(u, v)$ on S_0 (cf. 2.18).

(iii) By 3.3.2, ${}_n^nT_m^m\in K_3(S^0_{1/n+1/m})$ for $1/n+1/m<1/2$.

(iv) For every closed set $F\subset {}_nS^0$ there exists a closed oriented rectangle R such that $F\subset R\subset {}_nS^0$. By 2.7, ${}_m^my(u, v)\to {}_my$ uniformly to $y(u, v)$ on R. To show that ${}^nJ^m$ converges strongly to ${}^nJ$ on F, we shall verify that ${}^nJ^m\to_m {}^nJ$ a.e. on R and that $\{{}^nJ^m\}$ has property $(\mathcal{U})$ on R for fixed n; hence, by 2.3.1, $\iint_R |{}^nJ^m-{}^nJ|\,dudv\to_m 0$.

It follows directly from 2.19.1 that ${}^nJ^m\to_m {}^nJ$ a.e. on R. To show property $(\mathcal{U})$ for $\{{}^nJ^m\}$ on R, it suffices to consider the product

$$ \frac{\partial_n^nx(u, v)}{\partial u}\,\frac{\partial_m^my(u, v)}{\partial v} = \frac{\partial_n^nx(u, v)}{\partial u}\;{}_m^m\!\left(\frac{\partial y(u, v)}{\partial v}\right) $$

since the other product of ${}^nJ^m$ may be handled in the same way. By 2.8, the first factor is bounded for fixed n and, by 2.10, $\{{}_m^m(\partial y(u, v)/\partial v)\}$ has property $(\mathcal{U})$ on R; hence, from 2.2.2, $\{\partial_n^nx(u, v)/\partial u\cdot\partial_m^my(u, v)/\partial v\}$ has property $(\mathcal{U})$ on R for fixed n.

The conditions of the Closure Theorem being satisfied, we conclude that ${}_n^nT\in K_3({}_nS^0)$.

Exchanging the rôles of $x(u, v)$ and $y(u, v)$, we have the lemma:

3.4.1. LEMMA. *If T satisfies H_0, the transformation*

$$ T_n^n: \quad x = x(u, v), \qquad y = {}_n^ny(u, v), \qquad\qquad n > 2, (u, v) \in {}_nS, $$

belongs to class K_3 on ${}_nS^0$.

3.5. Our investigation of the transformations considered by Young will proceed according to the following scheme: First we approximate to the given transformation T by a suitable one of the transformations ${}_n^nT_n^n$, ${}_n^nT$, T_n^n. Then, as the reader will easily verify, all of the conditions of the Closure Theorem are satisfied except strong convergence of the Jacobians. Since the Jacobians of the approximating transformations converge a.e. to $J(u, v; T)$ (cf. 2.19.1), a necessary and sufficient condition for strong convergence is that the Jacobians of the approximating transformations have property $(\mathcal{U})$ (cf. 2.2, 2.3.1). We shall now investigate property $(\mathcal{U})$ for the Jacobians of the approximating transformations.

3.6. THEOREM 1 (see Young [2, pp. 80–85], cf. 1.9). *If T satisfies H_0 and one factor in each product $x_u(u, v)\cdot y_v(u, v)$, $x_v(u, v)\cdot y_u(u, v)$ is bounded on S_0, then $T\in K_3(S_0^0)$ and formula* (1) *holds whenever the integral on the left exists.*

Proof. Consider any fixed $R\subset S_0^0$. Choose n large enough that $R\subset {}_nS$. Approximate to T by ${}_n^nT_n^n$. To show property $(\mathcal{U})$ for $\{{}^nJ^n\}$ on R, it is sufficient to consider the product

$$\frac{\partial_n^n x(u, v)}{\partial u}\,\frac{\partial_n^n y(u, v)}{\partial v} = {}_n^n(x_u(u, v))\cdot {}_n^n(y_v(u, v))$$

since the other product in ${}^nJ^n$ may be handled in the same way. By assumption, one factor of $x_u(u, v)\cdot y_v(u, v)$, say $y_v(u, v)$, is bounded on S_0; hence, ${}_n^n(y_v(u, v))$ is uniformly bounded on R. By 2.10, $\{{}_n^n(x_u(u, v))\}$ has property $(\mathcal{U})$ on R. Therefore, from 2.2.2, $\{\partial_n^n x(u, v)/\partial u\cdot\partial_n^n y(u, v)/\partial v\}$ has property $(\mathcal{U})$ on R. By the Closure Theorem, $T\in K_3(S_0^0)$ and, by the result of Radó and Reichelderfer mentioned in 1.2, the theorem follows.

3.7. THEOREM 2 (see Young [2, pp. 80–85], cf. 1.9). *If T satisfies H_0 and in each product $x_u(u, v)\cdot y_v(u, v)$, $x_v(u, v)\cdot y_u(u, v)$, the factors belong to associated Lebesgue classes on S_0, then $T\in K_3(S_0^0)$ and formula* (1) *holds providing the integral on the left exists.*

Proof. Consider any fixed $R\subset S_0^0$. Choose n large enough that $R\subset {}_nS$. Approximate to T by ${}_n^nT_n^n$. On R

$$|{}_n^nJ_n^n| \leqq |{}_n^n(x_u(u, v))\cdot {}_n^n(y_v(u, v))| + |{}_n^n(x_v(u, v))\cdot {}_n^n(y_u(u, v))|.$$

It now follows from 2.11 and 2.2.1 that $\{{}^nJ^n\}$ has property $(\mathcal{U})$ on R. By the Closure Theorem, $T\in K_3(S_0^0)$; hence, the theorem follows.

3.8. THEOREM 3 (see Young [3, p. 163], cf. 1.9). *If T satisfies H_0; $|x_u(u, v)|$, $|y_u(u, v)|\leqq\lambda(u)$ and $|x_v(u, v)|$, $|y_v(u, v)|\leqq\mu(v)$, where $\lambda(u)$ and $\mu(v)$ are summable on (0, 1), then $T\in K_3(S_0^0)$ and formula* (1) *holds as soon as the integral on the left exists.*

Proof. Consider any fixed $R \subset S_0^0$. Choose n large enough that $R \subset {}_nS$. Approximate to T by ${}_nT$. To show property $(\mathcal{U})$ for $\{{}^nJ\}$ on R, it suffices to consider the product $\partial_n^n x(u, v)/\partial u \cdot \partial y(u, v)/\partial v$ since the other product of ${}^nJ$ may be handled in the same way. Let e be any measurable set comprised in R. Then

$$\iint_e \left|\frac{\partial_n^n x(u, v)}{\partial u}\right| \cdot \left|\frac{\partial y(u, v)}{\partial v}\right| du dv$$

$$\leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n}\left\{\iint_e |x_u(u+\alpha, v+\beta)| \cdot |y_v(u, v)|\, du dv\right\} d\alpha d\beta$$

$$\leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n}\left\{\iint_e \lambda(u+\alpha)\mu(v) du dv\right\} d\alpha d\beta$$

$$\leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n}\left\{\iint_{e_\alpha} \lambda(u)\mu(v) du dv\right\} d\alpha d\beta$$

$$\leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n} \Omega(\lambda(u)\mu(v), S_0, |e|)\, d\alpha d\beta$$

$$\leqq \Omega(\lambda(u)\mu(v), S_0, |e|).$$

Ω has been defined in 2.10 and

$$e_\alpha = \underset{(u,v)}{E}[(u-\alpha, v) \in e]$$

By the Theorem of Tonelli, $\lambda(u)\mu(v)$ is summable on S_0; hence, $\Omega(\lambda(u)\mu(v), S_0, |e|) \to_{|e|\to 0} 0$. Thus, by the last inequality and 2.2.1, $\{\partial_n^n x(u, v)/\partial u \cdot \partial y(u, v)/\partial v\}$ has property $(\mathcal{U})$ on R. Finally, by the Closure Theorem, $T \in K_3(S_0^0)$ and the theorem follows.

3.9. THEOREM 4 (cf. 3.1). *If T satisfies H_0; $|x_u(u, v)| \leqq \lambda(u)$, $|x_v(u, v)| \leqq \mu(v)$, where $\lambda(u)$ and $\mu(v)$ are summable on $(0, 1)$; and the total variations $V(v; y)$ and $V(u; y)$ are bounded on $(0, 1)$, then $T \in K_3(S_0^0)$ and formula (1) holds whenever the integral on the left exists.*

Proof. Consider any fixed $R \subset S_0^0$. Choose n large enough that $R \subset {}_nS$. Approximate to T by T_n. On R, $|J^n| \leqq \lambda(u) \cdot {}_n^n|y_v(u, v)| + \mu(v) \cdot {}_n^n|y_u(u, v)|$. This inequality, along with 2.14, 2.3, and 2.2.1, implies property $(\mathcal{U})$ for $\{J^n\}$ on R. By the Closure Theorem, $T \in K_3(S_0^0)$, so the theorem follows.

3.10. Young really does not discuss a situation corresponding to Theorem 4. Instead, he claims (see Young [2, p. 88]) to have established formula (3) of 1.4 under the conditions: $x(u, v)$ and $y(u, v)$ are absolutely continuous on every horizontal and on every vertical in S_0; $y_u(u, v)$, $y_v(u, v)$ are summable on S_0; $|x_u(u, v)| \leqq \lambda(u)$ and $|x_v(u, v)| \leqq \mu(v)$, where $\lambda(u)$, $\mu(v)$ are summable on $(0, 1)$ and $\lambda(u)y_v(u, v)$, $\mu(v)y_u(u, v)$ are summable on S_0. In attempting to prove this assertion, Young erroneously assumes that $y(u, v)$ can be written

as follows: $y(u, v) = y_1(u, v) - y_2(u, v)$, where $y_1(u, v)$ and $y_2(u, v)$ possess non-negative first partial derivatives in addition to all of the properties of $y(u, v)$ (see Young [2, p. 91]). Even if $y(u, v)$ could be split in this manner, $V(v; y)$ and $V(u; y)$ would be bounded (as we assumed in Theorem 4) since

$$V(v; y) \leqq V(v; y_1) + V(v; y_2) = \int_0^1 \frac{\partial y_1(u, v)}{\partial u} du + \int_0^1 \frac{\partial y_2(u, v)}{\partial u} du$$

$$\leqq y_1(1, v) - y_1(0, v) + y_2(1, v) - y_2(0, v)$$

and, similarly, $V(u; y) \leqq y_1(u, 1) - y_1(u, 0) + y_2(u, 1) - y_2(u, 0)$.

3.11. For the purpose of the next theorem, we require the following lemma:

LEMMA. *If $f(u, v) \geqq 0$ and $g(u, v) \geqq 0$ are summable on S_0 and if $0 < \delta < 1/2$ is fixed, then*

$$\phi(\alpha, \beta) = \iint_{S_\delta} f(u + \alpha, v + \beta) g(u, v) du dv$$

exists a.e. and is summable in r: $-\delta \leqq \alpha \leqq \delta$, $-\delta \leqq \beta \leqq \delta$.

Proof. See Young [2, p. 91].

3.12. THEOREM 5 (see Young [2, pp. 92–93]; cf. 1.9, 3.1). *For T satisfying H_0 and $0 < \delta < 1/2$ fixed, define*

$$T(\alpha, \beta; 0, 0): \quad x = x(u + \alpha, v + \beta), \qquad y = y(u, v), \qquad (\alpha, \beta) \in r, (u, v) \in S_\delta.$$

Then for a.e. (α, β) in r, $T(\alpha, \beta; 0, 0) \in K_3(S_\delta^0)$ and the transformation formula

$$\iint_{S_\delta} F(x(u + \alpha, v + \beta), y(u, v)) J(u, v; T(\alpha, \beta; 0, 0)) du dv$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(x, y) \cdot \mu(x, y; S_\delta, T(\alpha, \beta; 0, 0)) dx dy$$

holds if the integral on the left exists.

Proof. Consider any fixed $R \subset S_\delta^0$. Choose n_0 large enough that $R \subset S_{\delta+1/n}$ for $n > n_0$. Then for $(\alpha, \beta) \in r$ and $n > n_0$, define

$${}_n^n T(\alpha, \beta; 0, 0): \quad x = {}_n^n x(u + \alpha, v + \beta), \qquad y = y(u, v), \qquad (u, v) \in S_{\delta+1/n}.$$

We wish to show that $\{J(u, v; {}_n^n T(\alpha, \beta; 0, 0))\}$ has property $(\mathcal{U})$ on R for a.e. (α, β) in r. To this end, we need only consider the product $\partial {}_n^n x(u+\alpha, v+\beta)/\partial u \cdot \partial y(u, v)/\partial v$ since the other product of the Jacobian may be handled in the same way.

For $(\alpha, \beta) \in r$, define

$$G(u, v, \alpha, \beta) = |\, x_u(u + \alpha, v + \beta)\,| \cdot |\, y_v(u, v)\,|, \qquad (u, v) \in S_\delta,$$

and

$$G_n(u, v, \alpha, \beta) = {}_n^n|\, x_u(u + \alpha, v + \beta)\,| \cdot |\, y_v(u, v)\,|, \qquad n > n_0,\ (u, v) \in S_{\delta+1/n}.$$

By 3.11,

$$H(\alpha, \beta) = \iint_R G(u, v, \alpha, \beta)\,du\,dv$$

exists a.e. and is summable on r. Let E be the exceptional set of points (α, β) in r. From 2.9,

$$(1) \qquad {}_n^nH(\alpha, \beta) = \iint_R G_n(u, v, \alpha, \beta)\,du\,dv \underset{n}{\rightarrow} \iint_R G(u, v, \alpha, \beta)\,du\,dv = F(\alpha, \beta)$$

for a.e. $(\alpha, \beta) \in r$. Let E_2 be the exceptional set of points (α, β) in r. Consider a fixed point (α_0, β_0) in r and not belonging to $E_1 + E_2$. By 2.9, $G_n(u, v, \alpha_0, \beta_0) \rightarrow_n G(u, v, \alpha_0, \beta_0)$ a.e. on R; hence, it follows from (1) and 2.3 that $\{G_n(u, v, \alpha_0, \beta_0)\}$ has property $(\mathcal{U})$ on R. For $(u, v) \in R$ and $n > n_0$,

$$G_n(u, v, \alpha_0, \beta_0) \geqq \left|\frac{\partial_n^n x(u + \alpha_0, v + \beta_0)}{\partial u}\right| \cdot \left|\frac{\partial y(u, v)}{\partial v}\right|;$$

therefore, $\{\partial_n^n x(u+\alpha_0,\ v+\beta_0)/\partial u \cdot \partial y(u,\ v)/\partial v\}$ has property $(\mathcal{U})$ on R. Finally, by the Closure Theorem, $T(\alpha, \beta; 0, 0)$ belongs to class K_δ on S_δ^0 for a.e. (α, β) in r.

To complete the theorem, we must show that for a.e. (α, β) in r the image of the boundary of S_δ under $T(\alpha, \beta; 0, 0)$ is a set of planar measure zero. Since $x(u, v)$ is A.C.T. on S_0, for a.e. (α, β) in r the function $x(u+\alpha, v+\beta)$ is absolutely continuous on the boundary of S_δ; hence, for such (α, β), the continuous image of the boundary of S_δ under $T(\alpha, \beta; 0, 0)$ is a semi-rectifiable curve and its locus is a set of planar measure zero.

(For the proof of Theorem 5 we need not assume that the image of the boundary of S_0 under T is a set of planar measure zero.)

3.13. To prove the final theorem, we approximate to the given transformation T by the transformation

$${}_nT_n\colon \quad x = {}_nx(u, v), \qquad y = {}_ny(u, v), \qquad n > 2,\ (u, v) \in R_{1/n\ 0}.$$

Thus, to utilize the Closure Theorem, we must establish the lemma:

LEMMA. *If T satisfies H_0, the transformation ${}_nT_n$ belongs to class K_δ on $R_{1/n\ 0}^0$.*

PROOF. By 2.22, ${}_nx(u, v)$ and ${}_ny(u, v)$ are A.C.T. on $R_{1/n\ 0}$. From 2.15, $\partial_n x(u, v)/\partial u$ and $\partial_n y(u, v)/\partial u$ are continuous and, therefore, bounded on $R_{1/n\ 0}$. The result now follows from reasoning similar to that employed in 3.6.

3.14. THEOREM 6 (see Young [1, pp. 365–366; 2, p. 87]; cf. 1.9, 3.1). *If T satisfies H_0 and $|x_v(u, v)|$, $|y_v(y, v)| \leqq \mu(v)$, where $\mu(v)x_u(u, v)$ and $\mu(v)y_u(u, v)$ are summable on S_0, then $T \in K_3(S_0^0)$ and formula* (1) *holds providing the integral on the left exists.*

Proof. Consider any fixed $R \subset S_0^0$. Choose n large enough that $R \subset R_{1/n\ 0}$. Approximate to T by ${}_nT_n$. To show property $(\mathcal{U})$ for $\{J(u, v; {}_nT_n)\}$ on R, it suffices to consider the product $\partial_n x(u, v)/\partial u \cdot \partial_n y(u, v)/\partial v$ since the other product of the Jacobian may be handled in the same way. Let e be any measurable set comprised in R. Then

$$\begin{aligned}
&\iint_e \left|\frac{\partial_n x(u, v)}{\partial u}\right| \cdot \left|\frac{\partial_n y(u, v)}{\partial v}\right| du dv \\
&\quad \leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n}\left\{\iint_e |\, x_u(u+\alpha, v)\,| \cdot |\, y_v(u+\beta, v)\,|\, du dv\right\} d\alpha d\beta \\
&\quad \leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n}\left\{\iint_e \mu(v)\,|\, x_u(u+\alpha, v)\,|\, du dv\right\} d\alpha d\beta \\
&\quad \leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n}\left\{\iint_{e_\alpha} \mu(v)\,|\, x_u(u, v)\,|\, du dv\right\} d\alpha d\beta \\
&\quad \leqq \frac{n^2}{4}\int_{-1/n}^{1/n}\int_{-1/n}^{1/n} \Omega(\mu(v)\,|\, x_u(u, v)\,|, S_0, |\, e\,|)\, d\alpha d\beta \\
&\quad \leqq \Omega(\mu(v)\,|\, x_u(u, v)\,|, S_0, |\, e\,|),
\end{aligned}$$

where $\Omega(\mu(v)|x_u(u, v)|, S_0, |e|) \to_{|e| \to 0} 0$. Thus, by the last inequality and 2.2.1, $\{\partial_n x(u, v)/\partial u \cdot \partial_n y(u, v)/\partial v\}$ has property $(\mathcal{U})$ on R.

From 2.20.1 and 2.21.1, $J(u, v; {}_nT_n) \to_n J(u, v; T)$ a.e. on R; hence, as the reader can now easily verify, all of the conditions of the Closure Theorem are fulfilled. Therefore, $T \in K_3(S_0^0)$ and the theorem follows.

3.14.1. COROLLARY (see Young [3, pp. 189–190], cf. 1.9). *If T satisfies H_0; $|x_v(u, v)|$, $|y_v(u, v)| \leqq \mu(v)$, where $\mu(v)$ is summable on* (0, 1); *and the total variations $V(v; x)$ and $V(v; y)$ are bounded on* (0, 1), *then $T \in K_3(S_0^0)$ and formula* (1) *holds as soon as the integral on the left exists.*

Proof. Because of the Theorem of Tonelli, the relation

$$\int_0^1 \left\{\int_0^1 \mu(v)\,|\, x_u(u, v)\,|\, du\right\} dv = \int_0^1 \mu(v) V(v; x) dv$$

implies the summability of $\mu(v)x_u(u, v)$ on S_0, Likewise, $\mu(v)y_u(u, v)$ is summable on S_0.

3.15. The preceding six theorems account for all of Young's results on the

transformation of double integrals. Theorem 5 seems to offer a method of approach to the conjectured proposition (cf. 1.10):

If T satisfies H_0 and $J(u, v; T)$ is summable on S_0, then $T \in K_3(S_0^0)$ and formula (1) *holds if the integral on the left exists.*

Along this line, we have considered the question of deducing Theorems 1, 2, 3, 4, and 6 from Theorem 5. The first four theorems follow readily from the fifth, but we have not been able to deduce Theorem 6 from Theorem 5.

BIBLIOGRAPHY

H. E. BRAY
1. *Proof of a formula for an area*, Bull. Amer. Math. Soc. vol. 29 (1923) pp. 264–270.

G. FUBINI
1. *Sulla derivata seconda mista di un integrale doppio*, Rend. Cir. Mat. Palermo vol. 40 (1915) pp. 297– 98

R. G. HELSEL AND P. M. YOUNG
1. *Characterization theorems for integral means*, to appear soon in Duke Math. J.

E. J. MCSHANE
1. *Integrals over surfaces in parametric form*, Ann. of Math. (2) vol. 34 (1933) pp. 815–838.

C. B. MORREY
1. *A class of representations of manifolds.* I, Amer. J. Math. vol. 55 (1933) pp. 683–707.

H. RADEMACHER
1. *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln.* I, Math. Ann. vol. 79 (1919) pp. 340–359.

T. RADÓ
1. *On continuous transformations in the plane*, Fund. Math. vol. 27 (1936) pp. 201–211.
2. *A remark on the area of surfaces*, Amer. J. Math. vol. 58 (1936) pp. 598–606.
3. *On absolutely continuous transformations in the plane*, Duke Math. J. vol. 4 (1938) pp. 189–221.

T. RADÓ AND P. REICHELDERFER
1. *A theory of absolutely continuous transformations in the plane*, Trans. Amer. Math. Soc. vol. 49 (1941) pp. 258–307.

J. SCHAUDER
1. *Über stetige Abbildungen*, Fund. Math. vol. 12 (1928) pp. 47–74.

C. DE LA VALLÉE POUSSIN
1. *Sur l'intégrale de Lebesgue*, Trans. Amer. Math. Soc. vol. 16 (1915) pp. 435–501.

W. H. YOUNG
1. *On a formula for an area*, Proc. London Math. Soc. vol. 18 (1920) pp. 339–374.
2. *On a new set of conditions for a formula for an area*, Proc. London Math. Soc. vol. 21 (1923) pp. 75–94.
3. *Integration over the area of a curve and transformation of the variables in a multiple integral*, Proc. London Math. Soc. vol. 21 (1923) pp. 161–190.

THE OHIO STATE UNIVERSITY,
COLUMBUS, OHIO.

A THEORY OF CROSSED CHARACTERS

BY

REINHOLD BAER

The characters of the finite group(1) G in the cyclic group E of order(2) m are just the multiplicative, single-valued G to E functions. To define crossed characters we have to consider apart from G and E a homomorphism C of G into the group of automorphisms of E; and we denote the map of the element g in G under C also by g. A single-valued G to E function $f(g)$ is termed a C-character (or a crossed character) of G, if it is a solution of the following functional equation (3): $f(uv)=f(u)^v f(v)$ for u and v in G. The C-characters, like the ordinary characters of G, form a finite abelian group, the C-character group(4) of G; and as in the classical theory of characters one aims at establishing a duality between G and its C-character group. But here the similarity ends. For in the classical theory the following three properties are equivalent: (a) 1 is the only element in G which is mapped upon 1 by every character of G; (b) it is possible to establish a duality between G and its character group; (c) G and its character group are isomorphic. But if one substitutes in these properties C-characters for the ordinary characters, no such equivalence holds. Though still (c) implies (b) and (b) implies (a), only a few of the groups satisfying (a) meet the requirement (b); and only rarely a group satisfying (b) may be shown to satisfy (c).

Presented to the Society April 24, 1943; received by the editors December 21, 1942.

(1) Only *finite* groups will be considered in this investigation; for this reason we shall use the term "group" always in the sense of "finite group."

(2) It will be apparent immediately why it is not advisable to consider some "absolute" value group like the group of all the roots of unity.

(3) Apparently the first to investigate this functional equation was I. Schur. Recently functions meeting this requirement have been termed crossed characters, though it might have been more in conformity with accepted usage to term them "crossed homomorphisms" and to reserve the term "crossed character" for the crossed homomorphisms with values in a cyclic group. For investigations of this functional equation; see I. Schur, *Über die Darstellungen der endlichen Gruppen durch gebrochene lineare Substitutionen*, J. Reine Angew. Math. vol. 127 (1904) pp. 20–50; I. Schur, *Bemerkungen zu der vorstehenden Arbeit des Herrn Speiser*, Math. Zeit. vol. 5 (1919) pp. 7–10; A. Speiser, *Zahlentheoretische Sätze aus der Gruppentheorie*, Math. Zeit. vol. 5 (1919) pp. 1–6; R. Baer, *Automorphismen von Erweiterungsgruppen*, Actualités Scientifiques et Industrielles no. 205, 1935; S. MacLane and O. F. G. Schilling, *Normal algebraic number fields*, Trans. Amer. Math. Soc. vol. 50 (1941) pp. 295–384; A. H. Clifford and S. MacLane, *Factor sets of a group in its abstract unit group*, Trans. Amer. Math. Soc. vol. 50 (1941) pp. 385–406; S. MacLane and O. F. G. Schilling, *A general Kummer theory for function fields*, Duke Math. J. vol. 9 (1942) pp. 125–167.

(4) The C-character group of the Galois group of an algebraic extension is of importance in the theory of radical extensions as will be shown elsewhere.

In Chapter III of this investigation we characterize those groups G and homomorphisms C which meet the requirement (a); and those satisfying conditions (b) or (c) are determined in Chapter V. In Chapter II the order of the C-character group is evaluated and conditions for equality of the orders of G and of its C-character group are derived; this latter property is later seen to be intermediate between (b) and (c). Chapter IV is devoted to a characterization of the groups G and homomorphisms C satisfying the following postulate: the C-character group of G is the only group T of C-characters of G such that 1 is mapped upon 1 by every C-character in T. This property stands between (a) and (b) and is important by its applications to the theory of radical extensions of fields(5).

Chapter I. Preliminaries

I. 1. The classical theory of characters. In this section we give a resumé of the classical theory of characters of finite abelian groups in the form best suited to our purposes; and we indicate how to derive these statements from the theorems customarily found in the various texts(6).

If G is a finite (not necessarily abelian) group, and if E is a cyclic group of order m, then a *character of G in E* is a single-valued G to E function $f(g)$ satisfying the functional equation $f(g)f(h)=f(gh)$ for g and h in G; in short, a character of G in E is a homomorphism of the group G into the cyclic group E. If f' and f'' are characters of G in E, then $f'(g)f''(g)$ is a character of G in E which is called the product $f'f''$ of these characters. It is readily seen that the system of all the characters of G in E is a finite abelian group, the character group of G in E.

Basic for many constructions of characters is the following *extension principle.*

Lemma I.1.1. (a) *If S is a normal subgroup of the group G, and if f is a character of the quotient group G/S in E, then there exists one and only one character of G in E which maps S upon 1 and which induces f in G/S.*

(b) *If G is an abelian group the orders of whose elements are divisors of m, if f is a character of the subgroup S of G in E, then there exists a character of G in E which induces f in S.*

We denote by G' *the commutator subgroup of* G and by G^m the subgroup of G which is generated by all the mth *powers of elements in* G.

Theorem I.1.2. *$G'G^m$ is the subgroup of all the elements in G which are mapped upon 1 by every character of G in E.*

To prove this theorem one notes first that the characters of G in E are

(5) This application of the present theory will be treated elsewhere.

(6) For example, E. Hecke, *Theorie der algebraischen Zahlen*, Leipzig, 1923, in particular §10.

homomorphisms of the group G into a cyclic group of order m, and that homomorphisms of this kind map both G' and G^m upon 1; and one deduces from Lemma I.1.1 that there exists to every element not in $G'G^m$ a character of G in E which does not map it upon 1.

THEOREM I.1.3. *If G is a finite group and E a cyclic group of order m, then each of the following properties implies the others:*

(1) *1 is the only element in G which is mapped upon 1 by every character of G in E.*

(2) *G is an abelian group the orders of whose elements are divisors of m.*

(3) *G and its character group in E are isomorphic groups.*

The equivalence of (1) and (2) is essentially a restatement of Theorem I.1.2. Property (2) is a consequence of (3), since the character group of G in E is abelian, and since the mth power of every character of G in E is 1. That (2) implies (3) is a classical theorem(7).

If S is a subgroup of G, then we denote by $K(S)$ the group of all the characters f of G in E which map S upon 1, that is, $K(S)=\{f(S)=1\}$; and if T is a group of characters, then we denote by $G(T)$ the set of all the elements g in G which are mapped upon 1 by every character in T, that is, $G(T)=\{T(g)=1\}$.

THEOREM I.1.4(8). *If G is a finite abelian group the orders of whose elements are divisors of m, if E is a cyclic group of order m, then*

(a) *$G(K(S))=S$ for every subgroup S of G;*

(b) *$K(G(T))=T$ for every group T of characters of G in E;*

(c) *mapping the subgroup S of G upon $K(S)$ constitutes a duality(9) of G upon the character group of G in E.*

REMARK. If T is a group of characters such that $G(T)=1$, then every character of G is contained in T.

I. 2. **Fundamental concepts.** We are concerned with a cyclic group E of order m and a homomorphism C of the finite group G into the group of automorphisms of E. We shall use the same symbol g for the element g in the group G and the automorphism which corresponds to g under C, or as we shall say, the automorphism of E induced by the element g in G. Every automorphism of E consists in raising every element in E to a fixed power which is prime to m and which is uniquely determined modulo m; this integer (modulo m) which corresponds to the automorphism induced by the element g in G also shall be denoted by g.

The elements in G which are mapped upon the identity automorphism of

(7) See, for example, Hecke, op. cit. p. 36.

(8) See, for example, Hecke, op. cit. pp. 36–38.

(9) A *duality* is a monotone decreasing, 1:1 correspondence between the elements in two partially ordered sets.

E by C form a normal subgroup G_C of G; and G/G_C is essentially the same as a group of automorphisms of the cyclic group E. This implies in particular the commutativity of G/G_C.

It will be convenient to denote by C_0 the homomorphism of G which maps every element in G upon the identity automorphism of E; clearly C_0 is characterized by $G_{C_0}=G$.

The cyclic group E of order $m=\prod_p p^{m(p)}$ is the direct product of (uniquely determined) cyclic groups $E(p)$ of prime power order $p^{m(p)}$. Every automorphism of E induces an automorphism in every $E(p)$. If p is some prime, then we denote by Cp the homomorphism of G which maps the element g in G upon the automorphism of $E(p)$ which the automorphism g of E induces in $E(p)$. It is readily seen that G_C *is the cross-cut of all the* G_{Cp}.

A C-character of G *(in* E*) is a single-valued* G *to* E *function* $f(g)$ *which satisfies the functional equation* $f(gh)=f(g)^h f(h)$ *for* g *and* h *in* G. Clearly the C_0-characters of G are just the ordinary characters of G in E which we discussed in I.1.

If f' and f'' are C-characters of G, then $f'(g)f''(g)$ may be verified to be a C-character too; and this C-character of G is termed the product $f'f''$. The set of all C-characters of G thus forms a finite abelian group, *the C-character group* $L=L(G, C)$ *of* G; and the orders of the C-characters of G are divisors of the order m of E.

The order of a C-character f of G is a power of the prime number p if, and only if, f is at the same time a Cp-character of G, since f is of order a power of p if, and only if, f maps G upon a subgroup of $E(p)$.

If v is any element in E, then $f(g)=v^{1-g}$ for g in G is a C-character of G, since $v^{1-gh}=v^{1-h}v^{h-gh}=(v^{1-g})^h v^{1-h}$. Such C-characters of G have been termed *unit or principal characters*; and the group of all these characters may be called the *principal genus*. It is a cyclic group which is generated by the unit character e^{1-g} for e a generator of the cyclic group E.

LEMMA I.2.1. *If* f *is a* C*-character of* G*, and if* $G(f)$ *is the set of all the elements* g *in* G *which satisfy* $f(g)=1$*, then* (a) $G(f)$ *is a subgroup of* G*, and* (b) $f(g)=f(h)$ *if, and only if,* gh^{-1} *is in* $G(f)$.

REMARK. It is easy to construct examples of C-characters f such that $G(f)$ is not a normal subgroup of G.

Proof. If g and h are both in $G(f)$, then $f(gh)=f(g)^h f(h)=1$ so that gh also belongs to $G(f)$, proving (a). From $1=f(1)=f(hh^{-1})=f(h)^{h^{-1}}f(h^{-1})$ one infers that $f(gh^{-1})=f(g)^{h^{-1}}f(h^{-1})=(f(g)f(h)^{-1})^{h^{-1}}$ and this makes contention (b) evident.

LEMMA I.2.2. *If the automorphism* a *of the group* G *has the property that* g *and* g^a *are mapped by* C *upon the same automorphism of* E*, then the function* $f^a(x)=f(x^a)$ *is a* C*-character of* G *whenever* $f(x)$ *is a* C*-character of* G*; and in this*

fashion the automorphism a of G induces an automorphism a of the C-character group of G.

REMARK. The inner automorphisms of G meet the requirement imposed upon the automorphism a in the lemma, since the group of automorphisms of the cyclic group E is abelian.

Proof. If f is a C-character of G, and if g and g^a induce the same automorphism of E, then $f^a(gh)=f(g^ah^a)=f(g^a)^{h^a}f(h^a)=f^a(g)^hf^a(h)$, as was to be shown.

If p is a prime number, then we denote by $H(p)=H(G, p)$ *the set of all the elements g in G which are mapped upon* 1 *by all the Cp-characters of G.* It is a fairly immediate consequence of Lemma I.2.2. that $H(p)$ is a normal subgroup of G, since with $f(x)$ every $f(g^{-1}xg)=f'(x)$ for g in G is a Cp-character of G.

The cross-cut H of all the $H(p)$ is just the set of all the elements g in G which are mapped upon 1 by every C-character of G, since every C-character of G is a product of Cp-characters of G, as the Cp-characters of G are exactly the C-characters of order a power of p.

The cross-cut $\overline{H}(p)$ of all the $H(q)$ for $q\neq p$ will prove important occasionally; this normal subgroup of G is readily seen to consist of all those elements g in G which are mapped upon elements in $E(p)$ by every C-character of G.

The determination of the subgroups $H(p)$ will be one of our most important problems (see III.2 below); in this direction we now prove the following statement.

LEMMA I.2.3. $H(p)\leqq G_{Cp}$ *for every prime number p.*

Proof. If the element g in G does not belong to G_{Cp}, then it induces in $E(p)$ an automorphism which does not leave every element invariant. Thus there exists in $E(p)$ an element v such that $v\neq v^g$. The C-character $f(x)=v^{1-g}$ is a Cp-character and satisfies $f(g)\neq 1$ so that g does not belong to $H(p)$.

COROLLARY I.2.4. $H\leqq G_C$.

This is an immediate consequence of Lemma I.2.3, since G_C is the cross-cut of the G_{Cp}, and since H is the cross-cut of the $H(p)$.

I. 3. *C*-complete groups and the extension of *C*-characters. The group G is termed *C-complete*, if 1 is the only element in G which is mapped upon 1 by every C-character of G, that is, if the subgroup H, introduced in I.2, is equal to 1.

THEOREM I.3.1. *The group G is C-complete if, and only if, the following three conditions are satisfied by G and C:*

(i) *G_C is abelian*;
(ii) *the orders of the elements in G_C are divisors of the order m of E*;
(iii) *every (C_0-)character of G_C in E is induced by a C-character of G.*

Proof. Every C-character of G induces in G_C an ordinary, that is, a C_0-character in E. If G is C-complete, then 1 is the only element in G_C mapped upon 1 by every C_0-character of G_C which is induced by C-characters of G. The necessity of conditions (i), (ii) is now a consequence of Theorem I.1.3 and the necessity of condition (iii) may be deduced from Theorem I.1.4.

Assume conversely that conditions (i) to (iii) are satisfied by G and C. It is a consequence of Corollary I.2.4 that only elements in G_C are mapped upon 1 by every C-character of G. If $g \neq 1$ is an element in G_C, then we infer from conditions (i), (ii) and Theorem I.1.4 the existence of a C_0-character d of G_C in E such that $d(g) \neq 1$; and there exists by (iii) a C-character f of G, inducing d in G_C, so that in particular $f(g) \neq 1$. Hence G is C-complete.

COROLLARY I.3.2. *If G is C-complete, then*
(iv) *$g^{-1}xg = x^g$ for x in G_C and g in G where the exponent g is the modulo m uniquely determined integer corresponding to the element g under C.*

Proof. If x and y are elements in G_C, and if g is an element in G, then g and xgy induce the same automorphism in E. Hence C-characters f of G satisfy $f(g^{-1}xgx^{-g}) = f(g^{-1})^{xgx^{-g}} f(x)^{gx^{-g}} f(g)^{x^{-g}} f(x^{-g}) = f(g^{-1})^g f(g) f(x)^g f(x^{-g}) = f(g^{-1}g) f(x)^g f(x)^{-g} = 1$, since f is a C_0-character on G_C, and since x and x^{-g} are in G_C. Thus $g^{-1}xgx^{-g}$ is mapped upon 1 by every C-character of G; and this implies $g^{-1}xgx^{-g} = 1$, since G is C-complete, as was to be shown.

REMARK. If S is a subgroup of G_C, and if the conditions (i), (ii), (iv) are satisfied by G and C, then *S is a normal subgroup of G.*

LEMMA I.3.3. *If the normal subgroup S of G is part of G_C, then the homomorphism C of G induces a homomorphism C of G/S; and every C-character of G/S is induced by one and only one C-character of G which maps S upon* 1.

The proof is obvious.

LEMMA I.3.4. *If the conditions* (i), (ii), (iv) *are satisfied by G and C, if S is a subgroup of G_C and T a subgroup of G such that $G = ST$, if the C-characters s and t of S and T respectively coincide on the cross-cut $S \cap T$ of S and T, then there exists one and only one C-character of G which induces s in S and t in T.*

Proof. Every element g in G has the form uv for u in S and v in T since S is a normal subgroup of $G = ST$. If u, u' are elements in S and v, v' elements in T such that $uv = u'v'$, then v and v' induce the same automorphism in E, $u^{-1}u' = vv'^{-1}$ is an element in $S \cap T$, and s is a C_0-character in S; and hence we find that

$$s(u)^{v}t(v) = s(u)^{v}t(vv'^{-1}v') = s(u)^{v}t(vv'^{-1})^{v'}t(v')$$
$$= s(u)^{v'}s(u^{-1}u')^{v'}t(v') = s(u')^{v'}t(v').$$

This shows that $f(uv)=s(u)^{v}t(v)$ for u in S and v in T is a single-valued G to E function.

If u, u' are in S and v, v' are in T, then

$$f(uvu'v') = f(uvu'v^{-1}vv') = f(uu'^{v^{-1}}vv') = s(uu'^{v^{-1}})^{vv'}t(vv')$$
$$= s(u)^{vv'}t(v)^{v'}s(u'^{v^{-1}})^{vv'}t(v')$$
$$= [s(u)^{v}t(v)]^{v'}s(u')^{v'}t(v') = f(uv)^{v'}f(u'v')$$
$$= f(uv)^{u'v'}f(u'v'),$$

proving that f is the desired C-character of G.

That f is the only C-character of G meeting our requirements is obvious; and likewise it is obvious that the condition of the equality of s and t on $S\cap T$ is indispensable.

If the conditions (i), (ii) and (iv) are satisfied by the group G and the homomorphism C, then there exists one and essentially only one smallest group M which contains G_C and which is a direct product of cyclic groups of order m. If x is an element in G, then there corresponds to x under C a modulo m uniquely determined integer x; and thus y^x is a well determined element in M, whenever x is in G and y is in M. There exists therefore one and essentially only one group V which contains G and M as subgroups, satisfies $V=MG$ and the rule $x^{-1}yx=y^x$ for x in G and y in M. This group V we shall always denote by MG; and we note that M is a normal subgroup of MG whose cross-cut with G is exactly G_C.

THEOREM I.3.5. *If the conditions* (i), (ii) *of Theorem* I.3.1 *and condition* (iv) *of Corollary* I.3.2 *are satisfied by the group G and the homomorphism C, then the following property of MG is a necessary and sufficient condition for C-completeness of G:*

(v) *There exists a subgroup R of MG such that $MG=MR$ and $M\cap R=1$.*

REMARK. The elements of the subgroup R figuring in (v) form a set of representatives of the cosets of MG modulo M. Borrowing a term from the theory of extensions one may therefore restate (v) as follows: *M splits G.*

Proof. Since $M\cap G=G_C$, there exists a uniquely determined homomorphism of MG into the group of automorphisms of E which coincides with C on G and which maps every element in M upon the identity automorphism. This extension of the homomorphism C we denote by C too; and we mention that $(MG)_C=M$.

(+) *G is C-complete if, and only if, MG is C-complete.*

If MG is C-complete, then G is C-complete, since every C-character of MG

induces a C-character in its subgroup G. If conversely G is C-complete, and if f_0 is a C_0-character of M in E, then f_0 induces in G_C a C_0-character f_1; and we infer from Theorem I.3.1 (iii) the existence of a C-character f_2 of G which induces f_1 in G_C. Since f_0 and f_2 coincide on the cross-cut G_C of M and G, since $M=(MG)_C$, and since conditions (i), (ii), (iv) are satisfied by MG and C, we deduce from Lemma I.3.4 the existence of a C-character f of MG which coincides with f_2 on G and with f_0 on M. Thus conditions (i), (ii), (iii) of Theorem I.3.1 are satisfied by MG and C; and MG is C-complete, if G is C-complete.

(++) *MG is C-complete if, and only if, condition* (v) *is satisfied by MG.*

If condition (v) is satisfied by MG, then there exists a subgroup R of MG such that $MG=MR$ and $M\cap R=1$. Since conditions (i), (ii) and (iv) are satisfied by MG and C, and since $(MG)_C=M$, we deduce from Lemma I.3.4 that every C_0-character of M in E is induced by one and only one C-character of MG which maps R upon 1. Thus conditions (i), (ii), (iii) are satisfied by MG and C; and we have shown that the C-completeness of MG is a consequence of condition (v).

For the proof of the converse we need two lemmas which will prove useful in later investigations. If g is an element in G and f a C-character of G, then we put:

$$(*) \qquad F_g(f) = f(g).$$

LEMMA I.3.5.1. *F_g is, for every g in G, an ordinary character of the C-character group of G in E.*

This is an immediate consequence of the definition of product of C-characters.

LEMMA I.3.5.2. *If G is C-complete, then $g'=g''$ is a necessary and sufficient condition for $F_{g'}=F_{g''}$.*

$F_{g'}=F_{g''}$ if, and only if, $f(g')=f(g'')$ for every C-character f of G. We deduce from Lemma I.2.1 that this is the case if, and only if, $f(g'g''^{-1})=1$ for every C-character f of G; and this last fact is equivalent to $g'=g''$, since G is C-complete.

We suppose now that MG be C-complete; and we denote by K the C-character group of MG, by L the group of ordinary characters of K in E. It is a consequence of Lemma I.3.5.1 that F_x for x in MG is an element in L. If r, s are elements in M, then $F_{rs}(f)=f(rs)=f(r)\cdot f(s)=f(r)f(s)=F_r(f)F_s(f)$ or $F_{rs}=F_rF_s$; and we infer from Lemma I.3.5.2 that F_y is an isomorphism of M upon the subgroup F_M of L, since MG is C-complete. Since E is of order m, it follows that $L^m=1$. Since M and therefore F_M is a direct product of cyclic groups of the maximum order m in L, it follows that F_M is a direct factor of L,

proving the existence of a subgroup L' of L such that L is the direct product of F_M and L'.

We select in every coset X of MG/M a representative $r(X)$, in particular $r(M)=1$. If A and B are cosets of MG/M, then $(A, B)=(rAB)^{-1}r(A)r(B)$ is an element in M. Considering that C maps all the elements in the coset X upon the same automorphism X of E, we find for every C-character f of MG:

$$\begin{aligned} F_{r(AB)}F_{(A,B)}(f) &= f(r(AB))f((A, B)) = f(r(A, B))^{(A,B)}f((A, B)) \\ &= f(r(AB)(A, B)) = f(r(A)r(B)) = f(r(A))^{B}f(r(B)) \\ &= F^{B}_{r(A)}F_{r(B)}(f). \end{aligned}$$

Since $F_{r(X)}$ is an element in the direct product L of F_M and L', there exist elements $c(X)$ in M and $c'(X)$ in L' such that $F_{r(X)}=F_{c(X)}c'(X)$. Since (A, B) is in M, since the cross-cut of L' and F_M is 1, we deduce now that $c'(AB)=c'(A)^{B}c'(B)$ and that $F_{c(AB)}F_{(A,B)}=F^{B}_{c(A)}F_{c(B)}$. Since we have shown before that F effects an isomorphism between M and F_M, we may infer from the last identity the following equation: $c(AB)(A, B)=c(A)^{B}c(B)$ for A, B in GM/M.

Now we introduce new representatives of the cosets of MG/M by the definition: $r'(X)=r(X)c(X)^{-1}$ for X in MG/M. Then

$$\begin{aligned} r'(AB) &= r(AB)c(AB)^{-1} = r(AB)(A, B)c(A)^{-B}c(B)^{-1} \\ &= r(A)r(B)c(A)^{-B}c(B)^{-1} = r(A)c(A)^{-1}r(B)c(B)^{-1} \\ &= r'(A)r'(B); \end{aligned}$$

and this shows that the representatives $r'(X)$ of MG/M form a subgroup of MG, that is (v) is satisfied by MG. This completes the proof of $(++)$; and our theorem is an immediate consequence of $(+)$ and $(++)$.

The criterion (v) for C-completeness has the disadvantage of not being an "interior" property of G, since it involves extensions of the group G. This makes this criterion hard to handle in applications and for this reason we shall attack the problem from a different angle in Chapter III, in particular Theorem III.3.3.

The last paragraph of the proof of Theorem I.3.5 as well as the criterion (v) contain a comparatively simple method for actual construction of C-complete groups. This may be indicated briefly: If G is a C-complete group, then G_C is an abelian group which we shall denote by B, and B is part of the group M which is a direct product of cyclic groups of order m. The quotient group G/G_C is a group A of automorphisms of the cyclic group E of order m and is at the same time a group of automorphisms of M, since we may identify every element x in A with an integer modulo m that is prime to m. To determine G completely as an extension of B by A, realizing the given automorphisms of B, we need only the "factor sets" (a, b) for a and b in A which had been discussed in the last paragraph of the proof of Theorem I.3.5, and criterion

(v) states—as has been pointed out during this proof—exactly that there exists a single-valued A to M function $c(x)$, satisfying:

$$(**) \qquad c(x)^{v}c(y) = c(xy)(x, y) \qquad \text{for } x \text{ and } y \text{ in } A.$$

Putting $h(x)=Bc(x)$ for x in A we obtain a single-valued A to M/B function satisfying $h(x)^{v}h(y)=h(xy)$ for x, y in A, that is a crossed homomorphism of A into M/B.

If $c'(x)$ is another solution of the functional equation (**), then the function $c''(x)=c'(x)c(x)^{-1}$ satisfies $c''(x)^{v}c''(x)=c''(xy)$, and is therefore a crossed homomorphism of A into M. Passing from the factor set (a, b) to an "associated" factor set $(a, b)'=b(y)^{-1}b(x)^{-v}b(xy)(x, y)$ for $b(z)$ in B clearly does not change the crossed homomorphisms of A into M/B at all.

That conversely every crossed homomorphism of A into M/B may be realized in the fashion just described is readily seen.

Chapter II. The C-character group

It is the object of this chapter to give a somewhat crude description of the group of all the C-characters of a group G. This discussion, however, will be sufficient for determining the order of the C-character group of G, in case G is C-complete.

II. 1. **The automorphisms of a cyclic group.** For the convenience of the reader we collect here a number of known facts concerning the automorphisms of finite cyclic groups([10]).

If E is a cyclic group of order $m\neq 0$, then the group of all the automorphisms of E is a finite abelian group. If g is an automorphism of E, then g determines and is determined by a modulo m uniquely determined number g. Automorphism and number g are related to each other by the fact that the automorphism g of E maps every element in E upon its gth power. A number belongs to an automorphism if, and only if, it is prime to m. This relation between automorphisms and numbers is an isomorphism between the group of automorphisms of E and the multiplicative group of integers prime to m modulo m.

It is sufficient to discuss the case of a primary group E. Then the order of E is a prime power $m=p^{\mu}$.

Case A. *p is odd.* Then the group of all the automorphisms of E is a cyclic group of order $p^{\mu-1}(p-1)$.

Each of the following properties of an automorphism g of E and of the corresponding integer g modulo m implies all the others:

(1) The automorphism g is not of order a power of p.

(2) 1 is the only element in E which is left invariant by g.

(3) The integer g is not congruent to 1 modulo p.

([10]) See, for example, Hecke, op. cit. §13.

It will be convenient to term the automorphisms of order a power of p *regular automorphisms of* E and those meeting the requirements (1) to (3) *irregular automorphisms of* E.

If g is an irregular automorphism of E, then $g-1$ is prime to p and the order n of the automorphism g is the least positive integer k such that $(g^k-1)/(g-1)\equiv 0$ modulo m.

The (regular) automorphism g is of order $p^{\mu-j}$ for $0<j<\mu$ if, and only if, the integer g is congruent to $1+p^jg'$ modulo m for g' an integer prime to p. Then $(g^{p^{\mu-j}}-1)/(g-1)$ is divisible by $p^{\mu-j}$, but not by $p^{\mu-j+1}$.

Case B. $p=2$. If E is of orders 1 or 2, then the identity is the only automorphism of E. Thus we assume that $1<\mu$. Then E contains one and only one element of order 2 and this is left invariant by every automorphism of E.

The automorphism of E which maps every element in E upon its inverse shall be termed the *inversion of* E and shall be denoted by z. The group of all the automorphisms of E is the direct product of a cyclic group of order $2^{\mu-2}$ and of the group of order 2 which is generated by the inversion z.

The following two properties of an automorphism g of E and of the corresponding integer g modulo m are equivalent:

(i) The elements of orders 1 and 2 are the only elements in E which are left invariant by the automorphism g.

(ii) $g\equiv -1$ modulo 4.

We shall term the automorphisms of E meeting the requirements (i), (ii) *irregular automorphisms of* E, the others *regular automorphisms of* E. We note that the inversion is an irregular automorphism.

If the automorphism g of E is different both from the identity and the inversion, then the corresponding integer g is of the form $g\equiv \pm 1+2^jg'$ modulo m where g' is odd and $1<j<\mu$; and the order of the automorphism g is $2^{\mu-j}$.

If g is a regular automorphism of order $2^i\neq 1$, then $(g^{2^i}-1)/(g-1)$ is divisible by 2^i, though not by 2^{i+1}. If g is an irregular automorphism of order 2^i, not the inversion, then $(g^{2^i}-1)/(g-1)$ is divisible by $2^{\mu-1}$, but not by 2^μ.

We extend the definition of a regular automorphism of E, in case E is of orders 1 or 2, by saying that the identity automorphism is always regular.

The classification of automorphisms of E leads naturally to a *classification of the homomorphisms* C.

Suppose that E is a cyclic group of order $m=\prod_p p^{m(p)}$, $E(p)$ the subgroup of prime power order $p^{m(p)}$, that G is a finite group and that C is a homomorphism of G into the group of automorphisms of E. If every element in G induces (under C) in every $E(p)$ a regular automorphism, then C is termed *regular*, otherwise *irregular*. If in particular $1<m(2)$ and an element in G induces the inversion in $E(2)$, then C is said to be *singular* (so that a singular C is always irregular). We note that every G/G_{C_p} is of order a power of p, if C is regular, though this condition is not sufficient for regularity, since G/G_{C_2} is always of order a power of 2.

II. 2. **The C-characters of automorphism groups of primary cyclic groups.** In this section we are concerned with a complete determination of the C-characters of a group G which is mapped isomorphically by C upon a group of automorphisms of the cyclic group E of prime power order.

THEOREM II.2.1. *Suppose that C is a non-singular isomorphism of the group G into the group of automorphisms of the cyclic group E of prime power order $m=p^\mu$.*

(a) *Every C-character of G belongs to the principal genus so that the C-character group of G is a cyclic group.*

(b) *If C is regular, then G and its C-character group are of equal order.*

(c′) *If C is irregular and p is odd, then the C-character group of G is of order $m=p^\mu$.*

(c″) *If C is irregular and $p=2$, then the C-character group of G is of order $2^{\mu-1}$.*

Proof. We may assume throughout that $G\neq 1$.

Case A[11]. *p is odd.* Then the group of all the automorphisms of E is a cyclic group. Hence G is a cyclic group, generated by some element g.

Case A′. *C is regular.* Then the integer g corresponding to the automorphism g of E is of the form $g\equiv 1+p^jg'$ modulo m where g' is prime to p and $0<j<\mu$; and the order of g (and of G) is $p^{\mu-j}=n$.

If f is a C-character of G, then $1=f(g^n)=f(g)^s$ where the exponent $s\equiv 1+g+\cdots+g^{n-1}\equiv(g^{p^{\mu-j}}-1)/(g-1)$ is divisible by $p^{\mu-j}$, but by no higher power of p. Consequently $f(g)$ is an element in $E^{p^j}=E^{1-g}$. There exists therefore an element v in E such that $f(g)=v^{1-g}$ and it is readily verified that $f(g^i)=v^{1-g^i}$. Thus every C-character of G belongs to the principal genus; and G and its C-character group are of the same order $p^{\mu-j}$, since the unit character e^{1-x} for e a generator of E is of order $p^{\mu-j}$.

Case A″. *C is irregular.* Then neither G nor g is of order a power of p. Hence $g-1$ is prime to p so that $E=E^{1-g}$. There exists therefore to every C-character f of G an element v in E such that $f(g)=v^{1-g}$; and it is readily seen that $f(g^i)=v^{1-g^i}$. Thus every C-character of G belongs to the principal genus; and the order of the C-character group is m, since e^{1-x} for e a generator of E is of order m.

Case B. $p=2$. G contains just one element of order 2, since G does not contain the inversion of E. Hence G is cyclic (and $2<\mu$). Denote by g some element generating G.

Case B′. *C is regular.* Then $g\equiv 1+2^jg'$ modulo m where g' is odd and $1<j<\mu$; and an exact repetition of the argument used in Case A′ shows that every C-character of G belongs to the principal genus, and that G and its C-character group are of the same order $2^{\mu-j}$.

---

(11) These cases refer to the analogous cases in II.1.

Case B''. *C is irregular.* This occurs if, and only if, $g \equiv -1$ modulo 4, that is if, and only if, g is of the form $-1+2^j g'$ where g' is odd and $1<j<\mu$, since g is not the inversion of E. Then $1-g \equiv 2-2^j g' \equiv 2g''$ modulo m where g'' is odd. Hence $E^{1-g}=E^2$ showing that the order of the principal genus is $2^{\mu-1}$. The order of the automorphism g and of the group G is $2^{\mu-j}=n$. Thus every C-character f of G satisfies $1=f(g^n)=f(g)^s$ where the exponent $s=1+g+\cdots+g^{n-1}=(g^{2^{\mu-j}}-1)/(g-1)$ is divisible by $2^{\mu-1}$, but not by 2^μ. From this fact we deduce that $f(g)$ is an element in $E^2=E^{1-g}$; and it follows as usual that f belongs to the principal genus, completing the proof.

THEOREM II.2.2. *Suppose that C is a singular isomorphism of the group G into the group of automorphisms of the cyclic group E of order* $m=2^\mu$.

(a) *The principal genus is of order* $2^{\mu-1}$ *and the C-character group of G is of order* 2^μ.

(b') *If G is cyclic (of order 2), then the C-character group of G is cyclic.*

(b'') *If G is not cyclic, then the C-character group of G is the direct product of the principal genus and of the cyclic group of order 2 which consists of all the C-characters of G mapping the inversion z upon 1.*

Proof. We distinguish two cases.

Case 1. *G is cyclic.* Then G consists of the elements z and $z^2=1$, since z generates a direct factor of the group of all the automorphisms of E. If v is any element in E, then $f(z)=v$ defines a C-character of G because of $f(z)^z f(z)=1$. Hence the order of the C-character group of G is m. If u is the unit character e^{1-z}, then $u(z)=e^{1-z}=e^2$ so that the principal genus is of order $2^{\mu-1}$.

Case 2. *G is not cyclic.* Then there exists a basis of the abelian group G which contains z, and which therefore has the form z, g where g is an element of order $n=2^{\mu-j}$ with $1<j<\mu$.

If f is a C-character of G, then $f(g)^{-1}f(z)=f(g)^z f(z)=f(gz)=f(zg)=f(z)^g f(g)$ or

$$(+) \qquad f(g)^2 = f(z)^{1-g}.$$

Put $v=f(z)$. Then $v^{1-z}=v^2=f(z)^2$ and $v^{1-g}=f(z)^{1-g}=f(g)^2$; and $f(x)^2=v^{1-x}$ is a consequence of the fact that z and g generate G. Thus we have shown that the square of every C-character belongs to the principal genus.

It is a consequence of Theorem II.2.1 that every C-character of the cyclic group generated by g belongs to the principal genus. Hence there exists to every C-character f of G an element v in E such that $f(g)=v^{1-g}$; and we deduce from (+) that $v^{2(1-g)}=f(z)^{1-g}$. Since E is cyclic and g an automorphism, not the identity, this implies that $f(z)$ is in E^2. If w is an element in E such that $f(z)=w^2$, then the C-character $f'(x)=f(x)w^{x-1}$ has the property that $f'(z)=1$; and we have shown that every C-character f of G is modulo the principal genus congruent to a C-character of G which maps z upon 1.

Suppose now that r is a C-character of G such that $r(z)=1$. Then it follows

from $(+)$ that $r(g)^2=1$. If $r(g)=1$, then $r=1$. Since the element of order 2 in E is left invariant by every automorphism of E, there exists one and only one C-character of G (which is at the same time a C_0-character of G) which maps z upon 1 and g upon the element of order 2. Hence the principal genus is of index 2 in the C-character group of G. If the C-character r of G satisfies $r(z)=1\neq r(g)$, then r cannot be in the principal genus, since otherwise there existed an element s in E, satisfying $s^2=s^{1-z}=r(z)=1$ and $s^{1-g}=r(g)\neq 1$, an impossibility since g is odd and $1-g$ is even. Thus we have shown that the C-character group is the direct product of the principal genus and of a cyclic group of order 2. The order of the principal genus is 2^{a-1}, since $E^2=E^{1-z}$; and this completes the proof.

The following statement is an immediate consequence of the two preceding theorems.

COROLLARY II.2.3. *Suppose that C is an isomorphism of the group G into the group of automorphisms of the cyclic group E of prime power order $m=p^a$.*

(a) *The C-character group of G is cyclic if, and only if, G is cyclic.*

(b) *The C-character group of G is equal to the principal genus if, and only if, C is not singular.*

We conclude this section with a proof of the following fact.

COROLLARY II.2.4. *Suppose that C is an isomorphism of the group G into the group of automorphisms of the cyclic group E of prime power order $m=p^a$. Then the C-character f of G belongs to the principal genus if, and only if, one of the following conditions is satisfied:*

(1) *4 does not divide m;*

(2) *4 is a divisor of m, but none of the elements in G induces in E the inversion z;*

(3) *4 is a divisor of m, the group G contains an element z which induces in E the inversion z and $f(z)$ is in E^2, $f(g)=f(z)^{(1-g)2^{-1}}$ for every g in G.*

Proof. It is an immediate consequence of Corollary II.2.3(b) that every C-character of G is a principal character, if condition (1) or (2) is satisfied. Thus we assume now that neither (1) nor (2) be satisfied by G and C. If (3) is satisfied by the C-character f, then there exists an element v in E such that $f(z)=v^2$; and $f(g)=f(z)^{(1-g)2^{-1}}=v^{1-g}$ for every g in G is a principal character. If conversely f is a principal character, then there exists an element e in E such that $f(x)=e^{1-x}$ for every x in G. In particular $f(z)=e^{1-z}=e^2$ is in E^2 and $f(g)=e^{1-g}=f(z)^{(1-g)2^{-1}}$, since g is odd and $1-g$ is even.

II.3. **The principal genus.** It is obvious that C-characters in the principal genus map G_C upon 1. The problem to be discussed in this section is therefore that of characterizing the principal genus among the C-characters mapping G_C upon 1. We use the following *notations*:

(II.3.*) E is a cyclic group of order $m=\prod_p p^{m(p)}$, $E(p)$ the cyclic subgroup

of prime power order $p^{m(p)}$ of E; G is a group and C a homomorphism of G into the group of automorphisms of E.

THEOREM II.3.1. *The C-character f of G belongs to the principal genus if, and only if, the following conditions are satisfied by f:*

(1) *if the automorphism induced by the element g in G leaves every element in $E(p)$ invariant, then the order of $f(g)$ is prime to p;*

(2) *if 4 is a divisor of m, and if the element z in G induces the inversion in $E(2)$, then $f(z)$ belongs to E^2 and $f(g)^{m2^{-m(2)}} = f(z)^{(1-g)m2^{-m(2)-1}}$ for every g in G.*

Proof. If f is any C-character of G, then $f^{mp^{-m(p)}} = f_p$ is a Cp-character of G. If f is a principal character, then there exists an element v_p in $E(p)$ such that $f_p(x) = v_p^{1-x}$ is a principal Cp-character of G. Consequently $f_p(x) = 1$ for x in G_{Cp}, or equivalently, the order of $f(x)$ is prime to p for x in G_{Cp}. But x is in G_{Cp} if, and only if, the automorphism induced by x leaves every element in $E(p)$ invariant, showing the necessity of (1). The necessity of (2) is an immediate consequence of Corollary II.2.4 (3). If conversely conditions (1), (2) are satisfied by f, then we infer from (1) that f_p maps G_{Cp} upon 1; and hence it follows from (2) and Corollary II.2.4 (3) that f_p is a principal Cp-character of G. Thus there exists, for every prime p, an element v_p in $E(p)$ such that $f_p(x) = v_p^{1-x}$ for every x in G. Since $mp^{-m(p)}$ is prime to p, there exists an integer $i(p)$ such that $mp^{-m(p)}i(p) \equiv 1$ modulo $p^{m(p)}$. Put $v = \prod_p v_p^{i(p)}$. Then it is readily seen that $f(x) = v^{1-x}$ for every x in G, showing the sufficiency of the conditions (1), (2).

In the enunciation of the next theorem we make use of the fundamental subgroups $H(p)$, introduced in I.2. We note that $H(p)$ is exactly the set of all the elements in G which are mapped upon 1 by every Cp-character of G; and that $H(p) \leqq G_{Cp}$ by Lemma I.2.3.

THEOREM II.3.2[12]. *The principal genus consists of all the C-characters of G which map G_C upon 1 if, and only if,*

(1) *$H(p)G_C = G_{Cp}$ for every p, and*

(2) *C is not singular.*

REMARK. If E is of prime power order, and if G is C-complete, then condition (1) is automatically satisfied.

Proof. It has been pointed out before that $H(p) \leqq H(p)\, G_C \leqq G_{Cp}$; and it follows from the very definition of $H(p)$ that $G/H(p)$ is a Cp-complete group. Thus it follows from Theorem I.3.1 that $G_{Cp}/H(p)$ is an abelian group the orders of whose elements are divisors of $p^{m(p)}$, and that every C_0-character of $G_{Cp}/H(p)$ is induced by a Cp-character of $G/H(p)$. We note furthermore the obvious fact that every Cp-character of $G/H(p)$ is induced by one and only one Cp-character of G, since Cp-characters of G map $H(p)$ upon 1.

---

[12] Corollary III.2.3 below improves upon this theorem.

If $H(p)G_C < G_{Cp}$, then it follows from Theorem I.1.4 that there exists a C_0-character of $G_{Cp}/H(p)$ which maps $H(p)G_C$ upon 1, though it does not map every element in G_{Cp} upon 1; and consequently there would exist a Cp-character of G which does not map G_{Cp} upon 1, though it maps $H(p)G_C$ upon 1. It is obvious that such a C-character maps G_C upon 1 without belonging to the principal genus, proving the necessity of (1).

If $1 < m(2)$, and if there exists in G an element z inducing the inversion in $E(2)$, then we infer from Corollary II.2.3 (b) the existence of a $C2$-character of G which maps G_{C2} upon 1 and which does not belong to the principal genus, proving the necessity of (2).

Suppose conversely that conditions (1), (2) are satisfied by G; and that the C-character f of G maps G_C upon 1. Clearly f is the product of uniquely determined Cp-characters f_p, since a C-character is a Cp-character if, and only if, its order is a power of p. Since f_p is a suitable power of f, it follows that f_p maps G_C upon 1. Since $H(p)$ is mapped upon 1 by every Cp-character of G, we infer from (1) that G_{Cp} is mapped upon 1 by f_p. Now it follows from (2) and Corollary II.2.3(b) that every f_p belongs to the principal genus, showing that the product f of the f_p belongs to the principal genus too. The conditions (1), (2) are therefore sufficient.

COROLLARY II.3.3. *The principal genus consists of all the C-characters of G which map G_C upon 1, if the following two conditions are satisfied:*

(1) *if the automorphism g corresponding to the element g in G leaves the elements in $E(p)$ invariant, then its order is prime to p;*

(2) *C is not singular.*

Proof. Suppose that g is an element in G_{Cp}. Then it follows from (1) that its order modulo G_C is prime to p. Consequently $g = g'g''$ where the order of g' is prime to p, the order of g'' is a power of p, and where g'' belongs to G_C. Since g' is an element in G_{Cp} whose order is prime to p, and since Cp-characters of G induce C_0-characters in G_{Cp}, it follows that g' belongs to $H(p)$; and g belongs therefore to $H(p)G_C$, proving that $G_{Cp} = H(p)G_C$. Our contention is now an immediate consequence of Theorem II.3.2.

II.4. **Imbedding of C-complete groups into direct products.** It will be well to remember that C-complete groups may be characterized in two equivalent ways: a group G is C-complete if 1 is the only element mapped upon 1 by every C-character of G; and this is the case if, and only if, 1 is the cross-cut of the subgroups $H(p)$.

The following concepts will prove convenient for the enunciation of our imbedding theorem: suppose that $E(p)$ is a cyclic group of order $p^{m(p)}$ and that C_p is a homomorphism of the group G_p into the group of automorphisms of $E(p)$, where the primes p range over all the prime divisors of some integer m. We form the direct product E of the $E(p)$, the direct product G^* of

the groups G_p and we define the homomorphism C^* of G^* into the group of automorphisms of E as follows: if $\prod_p g_p$ for g_p in G_p is an element in G^*, then the automorphism of E corresponding under C^* to this element induces in $E(p)$ the same automorphism as the automorphism corresponding under C_p to g_p. This homomorphism C^* will always be referred to as *the direct product of the homomorphisms* C_p.

THEOREM II.4.1. *If C is a homomorphism of the group G into the group of automorphisms of E, and if G is C-complete, then G may be imbedded into the direct product of the quotient groups $G/H(p)$ in such a way that the direct product of the homomorphism Cp induces C in G.*

Proof. The homomorphism Cp of G into the group of automorphisms of $E(p)$ is well defined on $G_p=G/H(p)$, since $H(p)\leqq G_{Cp}$ by Lemma I.2.3. We denote by G^* the direct product of the groups G_p and by C^* the homomorphism of G^* into the group of automorphisms of E which we defined as the direct product of the homomorphisms Cp.

If g is an element in G, then put $g^*=\prod_p H(p)g$ an element in G^*; and it is readily seen that a homomorphism of G upon a subgroup of G^* is defined by mapping g upon g^*. If $g^*=1$, then g is contained in every $H(p)$ and $g=1$ is a consequence of the C-completeness of G. Thus G has been mapped isomorphically upon a subgroup of G^*. The automorphism of E corresponding under C^* to g^* induces in $E(p)$ the same automorphism which corresponds to g under Cp, that is the automorphism of E corresponding to g^* under C^* is the same as the automorphism of E which corresponds to g under C; and this completes the proof.

If we denote by $h(p)$ *the index of* $H(p)$ *in* G, that is the order of $G/H(p)$, then the following statement is an immediate consequence of Theorem II.4.1.

COROLLARY II.4.2. *If G is C-complete, then the l.c.m. of the $h(p)$ is a divisor of the order of G; and the order of G is a divisor of the product of the $h(p)$.*

We have introduced *the cross-cut* $\overline{H}(p)$ *of all the* $H(q)$ for $q\neq p$; this subgroup $\overline{H}(p)$ consists exactly of those elements in G which are mapped upon elements in $E(p)$ by every C-character of G. If G is C-complete, then the cross-cut of $H(p)$ and $\overline{H}(p)$ is 1 so that a coset of $G/H(p)$ contains at most one element in $\overline{H}(p)$.

THEOREM II.4.3. *If G is C-complete, then each of the following properties implies all the others:*

(1) *G is (essentially) the direct product of the $G/H(p)$.*

(2) *The order of G is $\prod_p h(p)$.*

(3) *Every coset of $G/H(p)$ contains an element in $\overline{H}(p)$.*

(4) *G is the direct product of the $\overline{H}(p)$; and $H(p)$ is the direct product of the $\overline{H}(q)$ for $q\neq p$.*

Proof. The equivalence of (1) and (2) is an obvious inference from Corollary II.4.2 (and Theorem II.4.1). To prove the equivalence of (1) and (3) we consider the map of G into the direct product G^* of the quotient groups $G/H(p)$ which we used in the proof of Theorem II.4.1. Clearly G is the direct product of the $G/H(p)$ if, and only if, G is mapped by this transformation upon the whole group G^*; and this is the case if, and only if, the following condition is satisfied:

(3′) If $C(p)$ is, for every prime p, a coset of $G/H(p)$, then the cross-cut of the sets $C(p)$ is not empty.

Suppose now that (3′) is satisfied; and that, for some given p, $U(p)$ is a coset of $G/H(p)$. Put $U(q)=H(q)$ for $q\neq p$. Then there exists by (3′) an element u contained in all these $U(r)$. Clearly u is in $\overline{H}(p)$, proving that (3) is a consequence of (3′) and therefore of (1).

Suppose conversely that (3) is satisfied, and that $C(p)$ is for every prime p (dividing m) a coset of $G/H(p)$. There exists by (3) to every p an element $c(p)$ in the cross-cut of $C(p)$ and $\overline{H}(p)$. Put $c=\prod_p c(p)$ (the order of the factors does not matter). If p is some prime, then $c\equiv c(p)$ modulo $H(p)$, since $c(q)$ for $q\neq p$ belongs to $\overline{H}(q)\leqq H(p)$; and this shows that c is an element in the cross-cut of the $C(p)$. Thus (3′) is a consequence of (3); and (3) implies (1), since (3′) implies (1).

If G is the direct product of the $\overline{H}(p)$, then it is clear that $H(p)$ is the direct product of the $\overline{H}(q)$ for $q\neq p$, and that G is, for every p, the direct product of $H(p)$ and $\overline{H}(p)$. But this last fact shows that (3) is a consequence of (4). If conversely (3) is satisfied by G, then G is the direct product of $H(p)$ and $\overline{H}(p)$, since in C-complete groups the cross-cut of $H(p)$ and $\overline{H}(p)$ is 1. Thus $\overline{H}(p)$ is of order $h(p)$ and the product of the subgroups $\overline{H}(p)$ is their direct product and is of order $\prod_p h(p)$, showing that G is, by Corollary II.4.2, the direct product of the $\overline{H}(p)$. Hence (3) and (4) are equivalent.

COROLLARY II.4.4. *If G is C-complete, and if $h(p)$ and $h(q)$ are relatively prime for $p\neq q$, then G is the direct product of the $G/H(p)$.*

REMARK. This criterion is satisfied, for example, whenever $h(p)$ is, for every p, a power of p.

Proof. If $h(p)$ and $h(q)$ are relatively prime for $p\neq q$, then the l.c.m. of the $h(p)$ is equal to their product. Hence we deduce from Corollary II.4.2 that $\prod_p h(p)$ is the order of G. Consequently we may infer from Theorem II.4.3 that G is the direct product of the $G/H(p)$.

II.5. **The order of the C-character group.** Noting that $H(p)\leqq G_{C_p}$ by Lemma I.2.3, we may decompose the index $h(p)$ into the product $h'(p)h''(p)$ of two indices where

$h'(p)=$*index of* $H(p)$ *in* G_{C_p}; and

$h''(p)=$*index of* G_{C_p} *in* G.

Furthermore *we define an invariant* $h'''(p)$ as follows:

$h'''(p) = h''(p)$, *if* Cp *is regular.*

$h'''(p) = p^{m(p)}$, *if* p *is odd and* Cp *is irregular.*

$h'''(2) = 2^{m(2)}$, *if* $C2$ *is singular.*

$h'''(2) = 2^{m(2)-1}$, *if* $C2$ *is neither regular nor singular.*

The motivation for this definition of $h'''(p)$ will be found in Theorems II.2.1 and II.2.2.

LEMMA II.5.1. $h'(p)$ *is a power of* p, *and* $h''(p) \leqq h'''(p)$.

Proof. Considering that, by the definition of $H(p)$, the group $G/H(p)$ is Cp-complete, it follows from Theorem I.3.1 that $h'(p)$ is a power of p. Since G/G_{Cp} is essentially a group of automorphisms of the cyclic group $E(p)$ of order $p^{m(p)}$, it follows that $h''(p)$ is a divisor of $p^{m(p)-1}(p-1)$, which makes $h''(p) \leqq h'''(p)$ evident.

THEOREM II.5.2[13]. *The order of the C-character group of* G *is* $\prod_p h'(p)h'''(p)$.

REMARK. If in particular E is of order a power of p, and if G is C-complete, then the order of the C-character group of G is $h'''(p)$ times the order of G_C.

Proof. Denote by $K(p)$ the group of those Cp-characters of G which map G_{Cp} upon 1. Then the order of $K(p)$ is, by Theorems II.2.1 and II.2.2, exactly $h'''(p)$, since every Cp-character of G/G_{Cp} is induced by one and only one Cp-character of G which maps G_{Cp} upon 1. It is a consequence of Theorem I.1.3 that the group of characters of $G_{Cp}/H(p)$ in $E(p)$ is exactly of order $h'(p)$. From Theorem I.3.1 we deduce that every character of $G_{Cp}/H(p)$ in $E(p)$ is induced by one, and therefore by $h'''(p)$, Cp-character of G, since $G/H(p)$ is Cp-complete. The group of Cp-characters of G is therefore of order $h'(p)h'''(p)$. But the C-character group of G is the direct product of the Cp-character groups of G, showing that its order is $\prod_p h'(p)h'''(p)$.

COROLLARY II.5.3. *Suppose that* G *is* C*-complete.*

(a) *The order of* G *does not exceed the order of the* C*-character group of* G.

(b) G *and its* C*-character group are of equal order if, and only if,* C *is regular.*

REMARK. If C is regular, then every $h''(p)$ is a power of p. Hence $h(p)$ is a power of p too, from Lemma II.5.1. Consequently we may deduce from Corollary II.4.4 that G is the direct product of the quotient groups $G/H(p)$.

Proof. It is a consequence of Corollary II.4.2 and of the C-completeness of G that the order of G is a divisor of $\prod_p h(p)$. We infer from Lemma II.5.1 that $h(p) \leqq h'(p)h'''(p)$. Hence (a) is a consequence of Theorem II.5.2.

---

[13] It may be deduced from Theorem III.2.2 (a) and Corollary III.3.2 below that $\prod_p h'(p)$ equals the order of G_C multiplied by $\prod_p h''''(p)$ where $h''''(p)$ is the index of $G_C H(p)$ in G_{Cp}.

If G and its C-character group are of equal order, then we deduce from (a) and Lemma II.5.1 that $h''(p)=h'''(p)$. This equality implies the regularity of C, as follows immediately from the definition of $h'''(p)$. If conversely C is regular, then we infer from Corollary II.4.4 and Theorem II.4.3 that $\prod_p h(p)$ is the order of G; and from the definition of $h'''(p)$ we infer that $h''(p)=h'''(p)$. Hence G and its C-character group are by Theorem II.5.2 of equal order, if C is regular.

Chapter III. Characterization of C-complete groups

III.1. **The case of primary** E. The following lemma will prove helpful in our discussion.

Lemma III.1.1. *If G is a cyclic group of order n, generated by the element g, if E is a cyclic group of order m, if C is a homomorphism of G into the group of automorphisms of E, and if v is an element in E, then $1=v^{1+v+\cdots+v^{n-1}}$ is a necessary and sufficient condition for the existence of a C-character f of G such that $f(g)=v$.*

Proof. The necessity of the condition is an immediate consequence of $1=f(1)=f(g^n)=f(g)^{1+v+\cdots+v^{n-1}}$. If the condition is satisfied, then a single valued G to E function f is defined by $f(g^i)=v^{1+\cdots+v^{i-1}}$ for $0<i\leqq n$; and this function satisfies $v=f(g)$, $1=f(g^n)=f(1)$. If $0<i, j\leqq n$, then there exists a uniquely determined integer $k=0, 1$ such that $0<i+j-kn\leqq n$; and we find that $f(g^ig^j)=f(g^{i+j-kn})=v^{1+\cdots+v^{i+j-kn-1}}=v^{1+\cdots+v^{i+j-1}}$, since, in case $k=1$, we may multiply by $1=v^{(1+\cdots+v^{n-1})v^{i+j-n}}$; and we find that

$$f(g^ig^j)=v^{1+\cdots+v^{i-1}v^i+\cdots+v^{i+j-1}}=v^{1+\cdots+v^{i-1}v(1+\cdots+v^{i-1})v^i}=f(g^i)f(g^i)^{v^i},$$

as was to be shown.

Theorem III.1.2. *If E is a cyclic group of prime power order $m=p^\mu$, and if C is a homomorphism of the group G into the group of automorphisms of E then the following four conditions are necessary and sufficient for C-completeness of G:*

(i) *G_C is abelian;*

(ii) *the orders of the elements in G_C are divisors of m;*

(iii) *$g^{-1}xg=x^v$ for g in G and x in G_C;*

(iv) *if $p=2$, and if the inversion of E is induced by the element z in G, then $z^2=1$.*

Proof. If G is C-complete, then we may infer conditions (i), (ii) from Theorem I.3.1, condition (iii) from Corollary I.3.2; and the necessity of (iv) may be seen as follows: if the element z in G maps every element in E upon its inverse, and if f is a C-character of G, then $f(z^2)=f(z)^zf(z)=f(z)^{-1}f(z)=1$; and the C-completeness of G implies $z^2=1$.

We suppose now that the conditions (i) to (iv) are satisfied by G, C.

Case I. *G/G_C is cyclic*[14]. Then there exists an element g which generates G modulo G_C. We denote by n' the order of G/G_C which is at the same time the order of g modulo G_C; and we put $g^*=g^{n'}$. If $\{y\}$ denotes the cyclic group, generated by the element y, then $\{g\}_C=\{g^*\}$.

The proof of the following statement will be the most important step.

(+) *There exists a C-character f of $\{g\}$ such that g^* and $f(g^*)$ are of equal order.*

We distinguish several cases.

Case 1. *n' is not a power of p.* Then p is odd, $g-1$ is prime to p; and it follows from (ii) that $x^g=x$ for x in G_C implies $x=1$. Hence it follows from (iii) that $g^*=g^{-1}g^*g=g^{*g}$ and that therefore $g^*=1$. (+) is trivially satisfied.

Case 2. *$p=2$ and g induces the inversion in E.* Then it follows from (iv) that $g^2=1$. Hence either $g^*=g^2=1$ and (+) is obvious; or else $g=g^*$ is an element in G_C; and the existence of the desired C-character which is a C_0-character is a consequence of (ii).

Case 3. *n' is a power of p; and in case $p=2$ the integer g corresponding to the element g under C is congruent to* 1 modulo 4. In this case the integer g is of the form $g\equiv 1+p^jg'$ modulo m where g' is prime to p, $0<j\leqq\mu$, and where

$$p = 2,\ 1 < \mu \quad \text{imply} \quad 1 < j.$$

Then we infer from the remarks in II.1 that $n'=p^{\mu-j}$ and that $1+g+\cdots+g^{n'-1}$ is divisible by $p^{\mu-j}$, but not by $p^{\mu-j+1}$. We note that n' is a divisor of the order n of g so that $n=n'n''$ where both n' and n'' are powers of p, since n'' is the order of the element g^* in G_C and since we may therefore apply (ii) on g^*. Since n' is the order of the automorphism g of E, it follows that $g^{n'}\equiv 1$ modulo m; and thus it follows that $n=n'n''$ is a divisor of $n''(1+g+\cdots+g^{n'-1})\equiv 1+g+\cdots+g^{n-1}$ modulo m. Finally we infer from condition (iii) that $g^*=g^{-1}g^*g=g^{*g}$ or $g\equiv 1+p^jg'\equiv 1$ modulo n'', since n'' is by (ii) a divisor of m; and this shows that n'' is a divisor of p^j, since n'' is a power of p whereas g' is prime to p. Hence $n=n'n''=p^{\mu-j}n''$ is a divisor of $p^\mu=m$.

From the last remark we infer the existence of an element v of order n in E. Since it has been shown that n is a divisor of $1+g+\cdots+g^{n-1}$ it follows that $v^{1+g+\cdots+g^{n-1}}=1$; and we infer from Lemma III.1.1 the existence of a C-character f of $\{g\}$ such that $f(g)=v$. Then $f(g^*)=f(g^{n'})=v^{1+g+\cdots+g^{n'-1}}$ is an element of order $n/n'=n''$, since $1+g+\cdots+g^{n'-1}$ is divisible by $n'=p^{\mu-j}$, but not by $n'p$. Since g^* is of order n'', the C-character f of $\{g\}$ meets all the requirements of (+).

Case 4. $p=2$, $1<\mu$, $g\equiv -1$ modulo 4, $g\not\equiv -1$ modulo m. In this case the integer g is of the form $g\equiv -1+2^jg'$ where g' is odd and $1<j<\mu$. Then we

[14] A proof of our theorem in Case I may be constructed which makes use of Theorem I.3.5.

infer from the remarks in II.1 that $n'=2^{\mu-i}$ and that $1+g+\cdots+g^{n'-1}$ is divisible by $2^{\mu-1}$, but not by 2^{μ}.

It is a consequence of (ii) and (iii) that $g^*=g^{-1}g^*g=g^{*g}$ or $2-2^ig'$ is divisible by the order n'' of g^*. Since n'' is a power of 2, it follows that $n''=1$ or 2, that is that $g^*=1$ or g^* is of order 2.

If $g^*=1$, then (+) is obvious. If g^* is of order 2, then $n=2n'$, $1+g+\cdots+g^{n-1}\equiv 2(1+g+\cdots+g^{n'-1})$ modulo m, since $g^{n'}\equiv 1$ modulo m. If v is any element of order m in E, then $1=v^{1+g+\cdots+g^{n-1}}$, since $1+g+\cdots+g^{n-1}$ is divisible by $2\ 2^{\mu-1}=m$; and it follows from Lemma III.1.1 that there exists a C-character f of $\{g\}$ such that $f(g)=v$. Then $f(g^*)=f(g^{n'})=v^{1+g+\cdots+g^{n'-1}}$ is of order 2, since the exponent is divisible by $2^{\mu-1}$, but not by 2^{μ}; and this shows that f meets all the requirements of (+).

Thus we have completed the proof of (+).

Suppose now that f_0 is a C_0-character of G_C in E. Then it follows from (+) that there exists a C-character f_1 of $\{g\}$ which coincides with f_0 on the cross-cut $\{g^*\}$ of G_C and $\{g\}$. Hence we may infer from Lemma I.3.4 the existence of a C-character f of G which induces f_0 in G_C and f_1 in $\{g\}$, since $G=G_C\{g\}$, and since the conditions (i) to (iii) are satisfied by G and C. This shows that every C_0-character of G_C in E is induced by a C-character of G. But this implies C-completeness of G, as follows from Theorem I.3.1, since (i), (ii) are satisfied by G and C.

Case II. *G/G_C is not cyclic.* In this case $p=2$; and G/G_C is the direct product of a cyclic group V by a cyclic group Z of order 2 such that the element of order 2 in Z maps every element in E upon its inverse (note that $2<\mu$). We denote by W the uniquely determined group between G_C and G which satisfies $W/G_C=V$; and we infer from what has been shown in Case I that W is C-complete.

If z is any element, representing the coset of order 2 in Z, then it follows from condition (iv) that z is of order 2. We note finally the obvious fact $G=W\{z\}$.

The proof of the following statement will be the most important step.

(++) *Every C-character of W is induced by a C-character of G.*

Mapping every element g in W upon the element zgz effects an automorphism of W; and g and zgz clearly induce the same automorphism of E. Hence it follows from Lemma I.2.2 that $f_1(x)=f(zxz)$ is a C-character of W whenever f is a C-character of W.

Suppose now that f is any C-character of W. Then it follows from the last remark that $f_2(x)=f(x)f_1(x)=f(x)f(zxz)$ is a C-character of W. If y is any element in G_C, then it follows from (iii) that $zyz=y^{-1}$; and since C-characters of W induce C_0-characters of G_C, we find that $f_2(y)=1$ for y in $G_C=W_C$. Since none of the automorphisms of E which are induced by elements in W maps every element in E upon its inverse, we may apply Theorem II.2.1 upon

f_2. Hence f_2 belongs to the principal genus of W, that is there exists an element v in E such that $v^{1-z}=f_2(x)=f(x)f(zxz)$ for x in W.

Since W is a normal subgroup of G, and since z is an element of order 2, generating G modulo W, it is possible to represent every element in G in one and only one way in the form $g=w(g)z^{i(g)}$ for $w(g)$ in W and $0\leqq i(g)\leqq 1$. Consequently a single-valued G to E function is defined by

$$f'(g) = f(w(g))^{(-1)^{i(g)}}v^{i(g)}$$

where v is the element in E introduced before.

The element g is in W if, and only if, $i(g)=0$ and $g=w(g)$, proving that f and f' coincide on W. If g and y are elements in G, then $gy=w(g)z^{i(g)}w(y)z^{i(y)}=w(g)z^{i(g)}w(y)z^{i(g)}z^{i(g)+i(y)}$. Thus $w(gy)=w(g)z^{i(g)}w(y)z^{i(g)}$ and $i(gy)\equiv i(g)+i(y)$ modulo 2, since W is a normal subgroup of G. From the choice of the element v it follows that $f(z^ixz^i)=v^{1-z^i}f(x)^{(-1)^i}$ for x in W and $i=0, 1$. This implies that

$$\begin{aligned} f[w(gy)] &= f[w(g)]^{w(y)}f[z^{i(g)}w(y)z^{i(g)}] \\ &= f[w(g)]^{w(y)}f[w(y)]^{(-1)^{i(g)}v^{1-w(y)^{i(g)}}}. \end{aligned}$$

One verifies furthermore by substituting the possible values for $i(g)$ and $i(y)$ that

$$[1 - w(y)^{i(g)}](-1)^{i(g)+i(y)} + i(gy) = i(y) + i(g)w(y)(-1)^{i(y)}.$$

Consequently we have finally

$$\begin{aligned} f'(gy) &= f[w(gy)]^{(-1)^{i(gy)}v^{i(gy)}} \\ &= f[w(g)]^{w(y)(-1)^{i(g)+i(y)}}f[w(y)]^{(-1)^{i(g)}v^{i(y)+i(g)w(y)(-1)^{i(y)}}} \\ &= \{f[w(g)]^{(-1)^{i(g)}v^{i(g)}}\}^{w(y)(-1)^{i(y)}}f[w(y)]^{(-1)^{i(y)}v^{i(y)}} \\ &= f'(g)^{v}f'(y). \end{aligned}$$

Thus f' has been shown to be a C-character of G which induces f in W; and this completes the proof of $(++)$.

Since W is C-complete and contains $G_C=W_C$, it follows from Theorem I.3.1 that every C_0-character of G_C in E is induced by a C-character of W; and hence we infer from $(++)$ that every C_0-character of G_C in E is induced by a C-character of G. The C-completeness of G is now an immediate inference of Theorem I.3.1 and conditions (i), (ii); and this completes the proof of the theorem.

We note that we have proved the following statement:

COROLLARY III.1.3. *If E is a cyclic group of order $m=2^\mu$, $1<\mu$, if C is a homomorphism of the group G into the group of automorphisms of E such that some elements in G are mapped by C upon the inversion of E, if G is C-complete*

(satisfies (i) *to* (iv) *of Theorem* III.1.2), *and if the subgroup W of G contains* G_C, *does not contain elements z inducing the inversion in E, though G is generated by adjoining z to W, then every C-character of W is induced by a C-character of G; and if f is a C-character of W, v an element in E, then there exists a C-character of G which induces f in W and maps z upon v if, and only if,* $v^{1-z}=f(x)f(zxz)$ *for x in W.*

III.2. **Determination of the subgroups** $H(p)$. Throughout this section we shall make use of the notations (II.3.*). For the enunciation of the next theorem it will be convenient to introduce the subgroup $H_1(p)=(G_{Cp})^{p^{m(p)}}(G_{Cp})'$ where X' is the commutator subgroup of the group X. Clearly $H_1(p)$ is the smallest subgroup of G_{Cp} such that $G_{Cp}/H_1(p)$ is an abelian group the orders of whose elements are divisors of $p^{m(p)}$ and $H_1(p)$ is a normal subgroup of G.

THEOREM III.2.1. (a) $H_1(p)\leqq H(p)\leqq G_{Cp}$.

(b) *If Cp is not singular, then* $H(p)/H_1(p)$ *is generated by the elements* $g^{-1}xgx^{-g}$ *for g in* $G/H_1(p)$ *and x in* $G_{Cp}/H_1(p)$.

(c) *If C2 is singular, then* $H(2)/H_1(2)$ *is generated by the elements* $g^{-1}xgx^{-g}$ *for g in* $G/H_1(2)$, *x in* $G_{C2}/H_1(2)$ *and by the squares of those elements in* $G/H_1(2)$ *which induce the inversion in* $E(2)$.

Proof. It is a consequence of the definition of $H(p)$ that $G/H(p)$ is Cp-complete. We infer $H(p)\leqq G_{Cp}$ from Lemma I.2.3; and we deduce from Theorem III.1.2 the validity of the following conditions:

(1) $G_{Cp}/H(p)$ is an abelian group the orders of whose elements are divisors of $p^{m(p)}$.

(2) $g^{-1}xg=x^g$ for g in $G/H(p)$ and x in $G_{Cp}/H(p)$.

(3) $z^2=1$, if $C2$ is singular, and if the element z in $G/H(2)$ induces the inversion in $E(2)$.

Condition (1) implies $H_1(p)\leqq H(p)$.

Now we define the subgroup $H_2(p)$ between $H_1(p)$ and $H(p)$ as follows: If Cp is not singular, then $H_2(p)/H_1(p)$ is generated by the elements $g^{-1}xgx^{-g}$ for g in $G/H_1(p)$ and x in $G_{Cp}/H_1(p)$. If, however, $C2$ is singular, then $H_2(2)/H_1(2)$ is generated by the elements $g^{-1}xgx^{-g}$ for g in $G/H_1(2)$, x in $G_{C2}/H_1(2)$ and by the squares of those elements in $G/H_1(2)$ which induce the inversion in $E(2)$. From conditions (2) and (3) it is immediately deduced that $H_2(p)\leqq H(p)$.

It is an immediate consequence of the definition of the subgroup $H_2(p)$ and of Theorem III.1.2 that the group $G/H_2(p)$ is Cp-complete. Thus it follows from Lemma I.3.3 that $H(p)\leqq H_2(p)$, proving the desired equality of $H(p)$ and $H_2(p)$.

We denote by $p^{J(p)}$ *the order of the subgroup of all those elements in* $E(p)$ *which are left invariant by every automorphism of* $E(p)$ *that is induced by an element in G.* Thus the integer corresponding to the element g in G under Cp

is of the form $g \equiv 1 + p^{J(p)} g_0$ modulo $p^{m(p)}$ for g_0 a suitable integer. From the facts enumerated in II.1 one readily deduces the following two statements:

$J(p) = 0$ if, and only if, $m(p) = 0$ or G/G_{Cp} is not of order a power of p.

$1 \leqq J(2)$ whenever $1 \leqq m(2)$.

For the investigations concerning the principal genus it is important to determine the subgroup $H^*(p)$ *of all those elements in* G *which are mapped upon* 1 *by all the* Cp*-characters of* G *mapping* G_C *upon* 1.

LEMMA III.2.2. (a) $H^*(p) = H(p)G_C$.

(b) *If* Cp *is not singular, then* $H^*(p)/G_C = (G_{Cp}/G_C)^{p^{J(p)}}$.

(c) *If* $C2$ *is singular, then* $H^*(2)/G_C$ *is generated by the elements in* $(G_{C2}/G_C)^2$ *and by the squares of those elements in* G/G_C *which induce the inversion in* $E(2)$.

We note that $J(2) = 1$, if $C2$ is singular.

Proof. It is an immediate consequence of the definition of $H^*(p)$ that $H(p)G_C \leqq H^*(p)$. Since G_C is mapped upon 1 by the principal genus, and since there exists to every element not in G_{Cp} a principal Cp-character which does not map it upon 1, it follows that $H^*(p) \leqq G_{Cp}$. If w is an element in G_{Cp}, though not in $H(p)G_C$, then we infer from the Cp-completeness of $G/H(p)$ and from Theorems I.3.1 and I.1.4 the existence of a Cp-character of G which maps $H(p)G_C$ upon 1, though it does not map w upon 1. Thus w cannot belong to $H^*(p)$ and hence $H^*(p) = H(p)G_C$.

G/G_C is essentially a group of automorphisms of a cyclic group. Hence G/G_C is an abelian group. It is an immediate consequence of the definition of $H^*(p)$ that $G/H^*(p)$ is Cp-complete; and thus it follows from Theorem III.1.2 that

(1) the orders of the elements in the abelian group $G_{Cp}/H^*(p)$ are divisors of $p^{m(p)}$,

(2) $x = g^{-1}xg = x^v$ for x in $G_{Cp}/H^*(p)$ and g in $G/H^*(p)$,

(3) $z^2 = 1$, if $1 < m(2)$ and if the element z in $G/H^*(2)$ induces the inversion in $E(2)$.

From (2) we infer that the orders of the elements in $G_{Cp}/H^*(p)$ are divisors of $p^{J(p)}$. Thus we define the subgroup $H^{**}(p)$ between G_C and G_{Cp} as follows: If Cp is not singular, then $H^{**}(p)/G_C = (G_{Cp}/G_C)^{p^{J(p)}}$. If $C2$ is singular, then $H^{**}(2)/G_C$ is generated by the squares of elements in G/G_C which induce in $E(2)$ the inversion or the identity. From what has been shown thus far we deduce $H^{**}(p) \leqq H^*(p)$. (The condition (1) may now be discarded, since $J(p) \leqq n_1(p)$.)

If x is an element in $G_{Cp}/H^{**}(p)$ and g an element in $G/H^{**}(p)$, then it follows from the construction of $H^{**}(p)$ that $G/H^{**}(p)$ is commutative, that the orders of the elements in $G_{Cp}/H^{**}(p)$ are divisors of $p^{J(p)}$, and that therefore $g^{-1}xg = x = x^v$ for x in $G_{Cp}/H^{**}(p)$ and g in $G/H^{**}(p)$. Hence by Theorem III.1.2 we have the Cp-completeness of $G/H^{**}(p)$, proving $H^*(p) = H^{**}(p)$.

THEOREM III.2.3. *The principal genus is exactly the set of the C-characters mapping G_C upon 1 if, and only if,*

(1) *C is not singular, and*

(2) $[G_{C_p}:G_C]$ *is prime to p for every prime p such that* $m(p)\neq 0$ *and* G/G_{C_p} *is of order a power of p.*

REMARK. $J(p)\neq 0$ if, and only if, $m(p)\neq 0$ and G/G_{C_p} is of order a power of p, as follows from the remarks in II.1.

Proof. Suppose first that the principal genus contains every C-character mapping G_C upon 1. Then it follows from Theorem II.3.2 that (1) is satisfied by G and that $G_{C_p}=H(p)G_C$. Hence we may deduce from Lemma III.2.2(a) that $G_{C_p}=H^*(p)$, and from Lemma III.2.2(b) that $G_{C_p}/G_C=H^*(p)/G_C=(G_{C_p}/G_C)^{p^{J(p)}}$. Consequently $J(p)\neq 0$ implies that the order of G_{C_p}/G_C is prime to p; and this shows the necessity of condition (2), since $J(p)\neq 0$ if, and only if, $m(p)\neq 0$ and G/G_{C_p} is of order a power of p.

If conversely conditions (1) and (2) are satisfied by G, it follows from Lemma III.2.2(a) and (b) and the above Remark that $(G_CH(p))/G_C=H^*(p)/G_C=(G_{C_p}/G_C)^{p^{J(p)}}=G_{C_p}/G_C$. Thus $G_CH(p)=G_{C_p}$; and we infer from Theorem II.3.2 that the principal genus contains all the C-characters, mapping G_C upon 1.

III.3. **The case of composite** E. We remind the reader of the fact that C-completeness of the group G is equivalent to the property: the cross-cut of all the subgroups $H(p)$ is 1. This latter property implies that the cross-cut of $H(p)$ and $\overline{H}(p)$ is 1, for every p, since $\overline{H}(p)$ is the cross-cut of all the $H(q)$ for $q\neq p$. The importance of the last criterion rests on the fact that $\overline{H}(p)$ consists of the elements mapped into $E(p)$ by every C-character of G.

THEOREM III.3.1. *Each of the following properties implies the others.*

(a) *G is C-complete.*

(b) *H(p) does not contain elements of order p in* G_C[15].

(c) G_C *is abelian and the cross-cut of H(p) and* G_C *consists of the elements of order prime to p in* G_C[15].

Proof. If y is an element of order p in G_C, then y belongs to $\overline{H}(p)$, since C-characters of G induce C_0-characters of G_C in E. If G is C-complete, then the cross-cut of $H(p)$ and $\overline{H}(p)$ is 1 so that elements of order p in G_C cannot be in $H(p)$. If G is not C-complete, then the cross-cut H of the subgroups $H(p)$ is different from 1 and is part of G_C so that there exists an element w of prime order r in H. This element w in G_C belongs to $H(r)$, since it belongs to H; and thus we have shown the equivalence of (a) and (b).

If G is C-complete, then G_C is abelian by Theorem I.3.1 so that G_C is the direct product of groups P and Q where P is of order a power of p and Q of

---

[15] Conditions (b) and (c) are supposed to be valid for every prime p.

order prime to p. It is a consequence of (b) that the cross-cut of $H(p)$ and P is 1; and it is a consequence of Theorem III.2.1 that Q is part of $H(p)$, that is Q is the cross-cut of G_C and $H(p)$. That (b) is a consequence of (c) is obvious; and this completes the proof.

COROLLARY III.3.2. *G is C-complete if, and only if, $H^*(p)$ is, for every prime p, the direct product of $H(p)$ and the p-component of the abelian subgroup G_C.*

Proof. If G is C-complete, then G_C is abelian by Theorem III.3.1(c). If G_C is abelian, then it is the direct product of two subgroups $U(p)$ and $V(p)$ where $U(p)$ consists of the elements of order a power of p in G_C and where $V(p)$ consists of the elements of order prime to p in G_C. Since C-characters of G induce C_0-characters of G_C, it follows that $V(p) \leqq H(p)$. Hence we deduce from Lemma III.2.2(a) that $H^*(p) = H(p)U(p)$. If G is C-complete, then it follows from Theorem III.3.1(b) that $H^*(p)$ is the direct product of $H(p)$ and $U(p)$. If conversely $H^*(p)$ is the direct product of $H(p)$ and $U(p)$, then the C-completeness of G is a consequence of Theorem III.3.1(b).

The criteria for C-completeness contained in Theorem III.3.1 and Corollary III.3.2, though sufficient for many applications, are unsatisfactory, since the subgroups $H(p)$, $H^*(p)$ which are defined in terms of C-characters are involved, whereas a satisfactory criterion should involve nothing but structural properties of G and C. We note that Theorem III.3.1(2) is really a family of conditions, namely one for every prime p. To solve our problem it is both sufficient and advantageous to find for every preassigned prime p, a criterion assuring the absence[16] of elements of order p in the cross-cut of $H(p)$ and G_C. The following concept will be used in the enunciation of such a criterion.

(III.3.*) $Q = Q(p)$ is the set of all the elements in G_{C_p} whose order is prime to p.

THEOREM III.3.3. *If G_C is abelian[17], and if p is a prime, then the following conditions are necessary and sufficient for the absence[16] of elements of order p in the cross-cut of $H(p)$ and G_C:*

(1) *$Q(p)$ is a normal subgroup of G and $G_{C_p}/Q(p)$ is an abelian group of order a power of p.*

(2) *If y is an element in G and x an element in G_{C_p} such that $xy^{-1}x^{-1}y$ is of order a power of p, then $(xy^{-1}x^{-1}y)^{p^{m(p)-J(p)}} = 1$; and if furthermore $x^{p^{J(p)}}$ is an element of order a power of p in G_C, then[18] $y^{-1}xy = x^v$.*

---

[16] This is property (b) of Theorem III.3.1, stated for one prime number p only.

[17] Theorem III.3.1 (c) shows that the commutativity of G_C is necessary for the completeness of the group G.

[18] Here the exponent y stands for a suitable integer which is modulo $p^{m(p)}$ congruent to the integer corresponding to the element y under C_p. It can be shown, however, that the order of the element x is a divisor of $p^{m(p)}$.

(3) *Suppose that* $p=2$, $1<m(2)$ *and that the element* y *in* G *induces the inversion in* $E(2)$[19].

(3′) *If* y^2 *is an element of order a power of* 2 *in* G_C, *then* $y^2=1$.

(3″) *If* g *is an element in* G *such that the integer corresponding to* g *under* $C2$ *is congruent to* -1 *modulo* 4, *if* x *is an element in* G_{C2} *such that* x, $g^{-1}xgx^{-1}$ *and* $y^{-1}xyx^{-1}$ *are of order a power of* 2, *then*[20] $g^{-1}xgx^{-1}=(y^{-1}xyx^{-1})^{(1-g)2^{-1}}$.

Proof. A. Assume that there are no elements of order p in the cross-cut of $H(p)$ and G_C.

We infer from Lemma I.2.3 that $H(p)\leqq G_{Cp}$; and from the definition of $H(p)$ that $G/H(p)$ is Cp-complete. Thus it follows from Theorem III.1.2 that:

(i) $G_{Cp}/H(p)$ is an abelian group the orders of whose elements are divisors of $p^{m(p)}$;

(ii) $v^{-1}xv=x^v$ for v in $G/H(p)$ and x in $G_{Cp}/H(p)$;

(iii) $z^2=1$ for z an element in $G/H(2)$ inducing the inversion in $E(2)$ provided $p=2$ and $1<m(2)$.

As an abelian group G_C is the direct product of subgroups P and K where P consists of all the elements of order a power of p in G_C and where K consists of all the elements of order prime to p in G_C. Since P and K are characteristic subgroups of G_C, they are normal subgroups of G. We infer from (i) that $K\leqq H(p)$; and from our hypothesis, that the cross-cut of $H(p)$ and P is 1. Hence K is the cross-cut of $H(p)$ and G_C.

From (i) and the commutativity of G/G_C it follows that the commutator subgroup G'_{Cp} of G_{Cp} is part of the cross-cut K of $H(p)$ and G_C. Hence G_{Cp}/K is abelian and is therefore the direct product of groups S and T where S consists of all the elements of order prime to p whereas T consists of all the elements of order a power of p. Denote by S^* the subgroup of G_{Cp} which contains K and satisfies $S=S^*/K$. The orders of the elements in S^* are prime to p, since this holds true for the orders of the elements both in K and in S^*/K, that is $S^*\leqq Q(p)$. If r is an element in $Q(p)$, then its order is prime to p and it is contained in a coset of G_{Cp}/K whose order is prime to p, that is r is in a coset belonging to S; and thus we have shown that $Q(p)=S^*$. Thus $Q(p)$ is a subgroup of G_{Cp}; it is therefore a characteristic subgroup of G_{Cp} and a normal subgroup of G. Since $K\leqq Q(p)$ and G_{Cp}/K is abelian, $G_{Cp}/Q(p)$ is abelian of order a power of p, proving the necessity of condition (1).

If y is any element in G, then we also denote by y some integer that corresponds modulo $p^{m(p)}$ to the element y under Cp. From the definition of $J(p)$ we infer the existence of an integer y_0 such that $y=1+p^{J(p)}y_0$.

If y is an element in G and x an element in G_{Cp}, then we infer from (ii) that

---

[19] This condition is vacuous, if none of the elements in G induces the inversion in $E(2)$.

[20] The exponent $(1-g)2^{-1}$ stands for any suitable integer meeting the obvious requirements; it is a consequence of the first part of condition (2) that the value of $(y^{-1}xyx^{-1})^{(1-g)2^{-1}}$ is independent of the particular choice of $(1-g)2^{-1}$.

$$y^{-1}xyx^{-v} = y^{-1}xyx^{-1}x^{1-v} = y^{-1}xyx^{-1}x^{-p^{J(p)}v_0}$$

is an element in $H(p)$. If both $y^{-1}xyx^{-1}$ and $x^{p^{J(p)}}$ are elements of order a power of p in G_C, then $y^{-1}xyx^{-v}$ is an element of order a power of p in the cross-cut of $H(p)$ and G_C so that $y^{-1}xyx^{-v}=1$, proving the necessity of the second part of condition (2). If we do assume only that $y^{-1}xyx^{-1}$ is an element of order a power of p (in G_C), then we infer from (i) and the above equation that

$$1 \equiv (y^{-1}xyx^{-v})^{p^{m(p)-J(p)}} \equiv (y^{-1}xyx^{-1}x^{-p^{J(p)}v_0})^{p^{m(p)-J(p)}}$$
$$\equiv (y^{-1}xyx^{-1})^{p^{m(p)-J(p)}}x^{-p^{m(p)}v_0} \equiv (y^{-1}xyx^{-1})^{p^{m(p)-J(p)}} \text{ modulo } H(p)$$

so that $(y^{-1}xyx^{-1})^{p^{m(p)-J(p)}}$ is an element of order a power of p in the cross-cut of $H(p)$ and G_C. But such an element is equal to 1, completing the proof of the necessity of (2).

Suppose now that $p=2$ and $1<m(2)$, that the element y in G induces the inversion in $E(2)$ and that y^2 is an element of order a power of 2 in G_C. Then it follows from (iii) that y^2 is an element of order a power of 2 in the cross-cut of G_C and $H(2)$, that is $y^2=1$, showing the necessity of (3′).

If $p=2$, $1<m(2)$, if y induces the inversion in $E(2)$ and if the integer corresponding to the element g in G under $C2$ is congruent to -1 modulo 4, then $(1-g)2^{-1}$ is an odd integer. If $g^{-1}xgx^{-1}$ and $y^{-1}xyx^{-1}$ for x an element in G_{C2} are elements of order a power of 2 (in G_C), then we infer from the commutativity of $G_{C2}/H(2)$ (see (i)!) that

$$g^{-1}xgx^{-1}(y^{-1}xyx^{-1})^{(g-1)2^{-1}} \equiv g^{-1}xgx^{-g}x^{g-1}(y^{-1}xyxx^{-2})^{(g-1)2^{-1}}$$
$$\equiv g^{-1}xgx^{-g}(y^{-1}xyx)^{(g-1)2^{-1}}x^{g-1}x^{-2(g-1)2^{-1}}$$
$$\equiv g^{-1}xgx^{-g}(y^{-1}xyx)^{(g-1)2^{-1}} \equiv 1 \text{ modulo } H(2)$$

by (ii); and thus $g^{-1}xgx^{-1}(y^{-1}xyx^{-1})^{(g-1)2^{-1}}=1$ as an element of order a power of 2 in the cross-cut of $H(2)$ and G_C, proving the necessity of (3″).

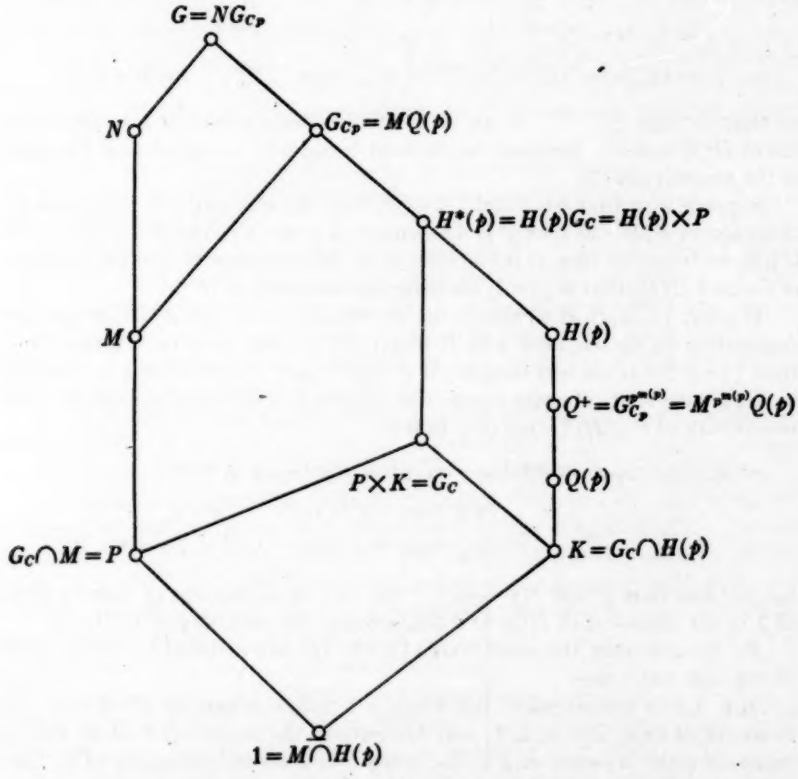
B. Assume that the conditions (1), (2), (3) are satisfied by G, C, p. We distinguish two cases.

B.1. *Cp is not singular.* We select a p-Sylow subgroup M of G_{Cp}. The cross-cut of $Q(p)$ and M is 1; and M contains the subgroup P of all the elements of order a power of p in G_C, since P is a normal subgroup of G. Since any two p-Sylow components of G_{Cp} are conjugate in G_{Cp}, and since every element in G transforms M into another p-Sylow subgroup of G_{Cp}, it follows that every coset of G/G_{Cp} contains elements from the normalizer of M in G. It follows from the special hypothesis B.1 that G/G_{Cp} is a cyclic group; and thus we have shown the existence of an element g in G which generates G modulo G_{Cp} and which transforms M into itself.

It is a consequence of condition (1) that $Q(p)$ is a normal subgroup of G, and that every coset of $G_{Cp}/Q(p)$ contains one and only one element in M.

Thus M and $G_{C_p}/Q(p)$ are isomorphic groups and we infer from (1) the commutativity of M.

If we denote by N the subgroup of G which is generated by adjoining the element g to M, then M is a normal subgroup of N, every coset of N/M is represented by some power of g, and all the elements in the same coset Mg^i of N/M induce in M the same automorphism.



We put $Q^+=Q^+(p)=G_{C_p}^{p^{m(p)}}$. From $Q(p)=Q(p)^{p^{m(p)}}$ and from the commutativity of M (together with the fact that every coset of $G_{C_p}/Q(p)$ contains one and only one element in M) we infer that

$$(') \qquad Q^+ = QM^{p^{m(p)}}.$$

Clearly it follows from (1) that $G_{C_p}/Q^+(p)$ is an abelian group the orders of whose elements are divisors of $p^{m(p)}$. It is readily seen now that $Q^+(p)$ is

exactly the group $H_1(p)$ introduced in Theorem III.2.1; and we infer from this theorem that $H(p)$ is the uniquely determined subgroup of G_{Cp} which contains Q^+ such that $H(p)/Q^+$ is exactly the subgroup of G_{Cp}/Q^+ which is generated by all the elements $v^{-1}wvw^{-v}$ for v in G/Q^+ and w in G_{Cp}/Q^+.

Since G_{Cp}/Q^+ is by (1) an abelian group the orders of whose elements are divisors of $p^{m(p)}$, it follows that every coset V of G/G_{Cp} induces a well-determined automorphism in G_{Cp}/Q^+ which we denote by V also. Thus $v^{-1}wvw^{-v} = w^{V-v}$ for v in G/Q^+, w in G_{Cp}/Q^+ and $V=G_{Cp}v$. From our choice of the element g we infer the existence of an integer i such that $V=G_{Cp}g^i$; and for the integers corresponding under Cp to these elements it follows likewise that $v\equiv g^i$ modulo $p^{m(p)}$. If we put $Z=G_{Cp}g$, then we find for v in G/Q^+ and w in G_{Cp}/Q^+:

$$v^{-1}wvw^{-v} = w^{V-v} = w^{Z^i-g^i} = w^{(Z^{i-1}+Z^{i-2}g+\cdots+Zg^{i-2}+g^{i-1})(Z-g)}.$$

From this equation one readily deduces the following fact:

(″) $$H(p)/Q^+ = (G_{Cp}/Q^+)^{Z-g}.$$

Now let g stand for any integer which modulo $p^{m(p)}$ corresponds under Cp to the element g in G. If x is an element in $H(p)$, then we infer from (″) the existence of an element Y in G_{Cp}/Q^+ such that $Q^+x=Q^+Y^{Z-g}$. But Y contains an element y in M. Hence it follows from (′) that x is in

$$Q^+(Q^+y)^{Z-g} = Q^+g^{-1}ygy^{-g} = QM^{p^{m(p)}}g^{-1}ygy^{-g}.$$

Thus we have proved the following statement.

(‴) Every element in $H(p)$ has the form:

$$x = x_0x_1^{p^{m(p)}}g^{-1}x_2gx_2^{-g} \quad \text{for} \quad x_0 \text{ in } Q(p),\ x_1 \text{ and } x_2 \text{ in } M.$$

We note that by the choice of g the element $g^{-1}x_2g$ is in M too so that $x_0^{-1}x$ belongs to M.

Suppose now that x is an element of order a power of p in the cross-cut of $H(p)$ and G_C. Then it follows from (‴) and the remark following (‴) that both x and $x_0^{-1}x$ belong to M, since $P\leqq M$. Hence x_0 is in the cross-cut of $Q(p)$ and M so that $x_0=1$. Since $g^{-1}x_2gx_2^{-1}$ is in the cross-cut P of M and G_C, it follows that $x(g^{-1}x_2gx_2^{-1})^{-1}=x_1^{p^{m(p)}}x_2^{1-g}$ is in P too. (Use the commutativity of M.) From the definition of the integer g and the invariant $J(p)$ we infer the existence of an integer g^* such that $1-g=g^*p^{J(p)}$; and since the element g generates G modulo G_{Cp}, that is the automorphism corresponding to g under Cp generates the group of automorphisms of $E(p)$ induced by elements in G, it follows that g^* is prime to p. Hence

$$(x_1^{p^{m(p)-J(p)}}x_2^{g^*})^{p^{J(p)}} = x_1^{p^{m(p)}}x_2^{1-g}$$

is in G_C and even in P; and it follows from condition (2) (second part) that

$$g^{-1}(x_1^{p^{m(p)-J(p)}g^*}x_2)g = (x_1^{p^{m(p)-J(p)}}x_2^{g^*})^g,$$
$$(g^{-1}x_2gx_2^{-g})^{g^*} = g^{-1}x_2^{g^*}gx_2^{-g^*g} = x_1^{p^{m(p)-J(p)}g}\,g^{-1}x_1^{-p^{m(p)-J(p)}}g;$$

and consequently we find that

$$\begin{aligned} x^{g^*} &= x_1^{g^*p^{m(p)}}(g^{-1}x_2gx_2^{-g})^{g^*} \\ &= x_1^{g^*p^{m(p)}}x_1^{p^{m(p)-J(p)}(1-g^*p^{J(p)})}(g^{-1}x_1g)^{-p^{m(p)-J(p)}} \\ &= x_1^{p^{m(p)-J(p)}}(g^{-1}x_1^{-1}g)^{p^{m(p)-J(p)}} = (x_1g^{-1}x_1^{-1}g)^{p^{m(p)-J(p)}} \\ &= 1 \end{aligned}$$

as follows from the first part of condition (3). But g^* is prime to p; and hence it follows that $x=1$; and we have shown that the only element of order a power of p in the cross-cut of $H(p)$ and G_C is 1, as we desired to do.

B.2. $p=2$ *and C2 is singular.* Then $J(2)=1$, as has been pointed out before.

We select a 2-Sylow subgroup N of G and denote by M the cross-cut of N and G_{C2}. Since G_{C2} is a normal subgroup of G, it follows that M is a 2-Sylow subgroup of G_{C2}. Since both $G_{C2}/Q(2)$ (by (1)) and G/G_{C2} are of order a power of 2 (see II.1.B), it follows that every coset of $G/Q(2)$ contains one and only one element in N; and we infer from (1) the commutativity of M, as M and $G_{C2}/Q(2)$ are isomorphic.

We denote by Y the set of all the elements in G which induce the inversion in $E(2)$. Clearly Y is a coset of G/G_{C2} and the cross-cut Y^* of Y and N is a coset of N/M.

We denote by $Q^+=Q^+(2)$ the subgroup $G_{C2}^{2^{m(2)}}$; and as in B.1 it follows that $Q^+(2)=Q(2)M^{2^{m(2)}}$ and that $G_{C2}/Q^+(2)$ is an abelian group the orders of whose elements are divisors of $2^{m(2)}$. Then one deduces from Theorem III.2.1 that

(+) $H(2)$ is the uniquely determined subgroup of G_{C2} which contains $Q^+(2)$ such that $H(2)/Q^+(2)$ is generated by the elements y^2 for y in Y and by the elements $v^{-1}wvw^{-v}$ for v in $G/Q^+(2)$ and for w in $G_{C2}/Q^+(2)$.

It will be convenient to start with the investigation of a (possibly smaller) subgroup, namely the subgroup Q^0 between $Q^+(2)$ and $H(2)$ which is uniquely determined by the property that Q^0/Q^+ is generated by the elements y^2 for y in the coset Y of G/G_{C2} [consisting of all the elements in G/Q^+ which induce the inversion in $E(2)$]. Noting that every coset of $G/Q(2)$ contains one and only one element in N and that the inverses of elements in Y belong to Y, it is readily verified that every element x in Q^0 has the form:

$$x = x_0 x_1^{2^{m(2)}} y_1^2 \cdots y_k^2 \qquad \text{for } 0 \leqq k,\ x_0 \text{ in } Q,\ x_1 \text{ in } M,\ y_i \text{ in } Y^*.$$

If $1<k$, then it is possible to simplify this normal form for elements in Q^0 as follows: Since every y_1 is an element in N which induces the inversion in $E(2)$, all the elements y_i are in the same coset Y^* of N/M; and hence there exist elements z_i in M such that $y_i = z_i y_1$. Since all the elements in Y^* induce in the abelian group M the same automorphism which we denote by Y^*, it follows that $y_i^2 = z_i^{1+Y^*} y_1^2$. Since z_i and y_1^2 are elements in the commutative group M, we deduce that $y_1^2 \cdots y_k^2 = (z_1 \cdots z_k)^{1+Y^*} y_1^{2k}$. Considering that an odd power of an element in Y^* is an element in Y^*, we see finally that every element x in Q^0 has the form:

$$(*) \qquad x = x_0 x_1^{2^{m(2)}} x_2^{1+Y^*} x_3^{2^h}$$

for x_0 in Q, x_1 and x_2 in M, where $x_3 = 1$ if $h = 0$ and x_3 is in Y^* if $0<h$.

Now we distinguish two cases:

B.2.1. *G/G_{C2} is of order* 2. Then the elements in G induce in $E(2)$ the inversion and the identity; and it follows from $(+)$ and $(*)$ that every element x in $H(2)$ has the form given in $(*)$. Suppose now that such an element x belongs to P (is an element of order a power of 2 in G_C). Since $P \leqq M$, since M is an abelian 2-group, and since $x_0^{-1}x$ and x both belong to M, it follows immediately that $x_0 = 1$. We note furthermore that $x_i^{Y^*-1}$ belongs to the cross-cut P of G_C and M, since it belongs to the commutator subgroup of G which is part of G_C.

If $h=0$, then $x = x_1^{2^{m(2)}} x_2^{1+Y^*}$ and consequently $(x_1^{2^{m(2)-1}} x_2)^{2^{J(2)}} = x_1^{2^{m(2)}} x_2^2 = x\, x_2^{1-Y^*}$ is an element in P. Hence it follows from the second part of condition (2) that $(x_1^{2^{m(2)-1}} x_2)^{Y^*} = (x_1^{2^{m(2)-1}} x_2)^{-1}$ or $x_2^{Y^*+1} = x_1^{-(Y^*+1)2^{m(2)-1}}$; and thus we find that

$$x = x_1^{2^{m(2)} - (Y^*+1)2^{m(2)-1}} = x_1^{(1-Y^*)2^{m(2)-J(2)}} = 1$$

by the first part of condition (2).

Next if $h=1$, then

$$x = x_1^{2^{m(2)}} x_2^{1+Y^*} x_3^2 = x_1^{2^{m(2)}} (x_2 x_3)^2 = x_1^{2^{m(2)-1}(1-Y^*)} x_1^{2^{m(2)-1}(1+Y^*)} (x_2 x_3)^2$$
$$= x_1^{2^{m(2)-1}(1-Y^*)} (x_1^{2^{m(2)-1}} x_2 x_3)^2$$

and this shows that $x_1^{2^{m(2)-1}} x_2 x_3$ is an element in Y^* whose square is in P. Thus it follows from condition $(3')$ that this square is 1 so that $x = x_1^{(1-Y^*)2^{m(2)-J(2)}} = 1$ by condition (2) (first part).

If finally $1<h$, then $2^{h-1}-1$ is odd; and we find that

$$x = x_1^{2^{m(2)-1}(1-Y^*)} (x_1^{2^{m(2)-1}} x_2 x_3)^2 x_3^{(2^{h-1}-1)2}.$$

Since both $x_3^{\pm(2^{h-1}-1)}$ and $x_1^{2^{m(2)-1}}x_2x_3$ are elements in Y^*, there exists an element z in M such that $x_1^{2^{m(2)-1}}x_2x_3=zx_3^{1-2^{h-1}}$; and consequently we have $x=x_1^{2^{m(2)-1}(1-Y^*)}z^{1+Y^*}$. The first of the factors is 1 as a consequence of the first part of condition (2); and hence $x=1$ is a consequence of the second part of (2). This completes the proof of the fact that in the case under consideration the cross-cut of P and $H(2)$ is 1.

B.2.2. *G/G_{C2} is not of order* 2. Then we infer from II.1.B and from the existence of elements in G which induce the inversion in $E(2)$ that there exists a basis Z, Y of G/G_{C2} where Y induces the inversion in $E(2)$ and where the integer corresponding to Z under $C2$ is congruent to -1 modulo 4. We note that the cross-cuts Z^* and Y^* of Z and Y respectively with N are cosets of N/M which induce the automorphisms Z^* and Y^* respectively in the commutative group M.

Now one derives from (+) and (*) with the same arguments as were used in B.1 that every element x in $H(2)$ has the form:

$$(**) \qquad x = x_0x_1^{2^{m(2)}}x_2^{1+Y^*}x_3^{Z^*-g}x_4^{2^h}$$

where x_0 is in $Q(2)$, where x_1, x_2 and x_3 are elements in M and x_4 is an element in Y^* if $0<h$, and $x_4=1$ if $0=h$, and where g is an integer which is modulo $2^{m(2)}$ congruent to the integers corresponding under $C2$ to the elements in Z^*.

If x is an element in the cross-cut of $H(2)$ and P, then it follows as usual that $x_0=1$. It follows from condition (3'') and the first part of condition (2) that $x_1^{(1-Y^*)2^{m(2)-1}}=1$ and that $x_3^{(Z^*-g)+(1+Y^*)(g-1)2^{-1}}=1$; and thus we find that

$$x = x_1^{2^{m(2)-1}(1+Y^*)}x_2^{1+Y^*}x_3^{(1+Y^*)(1-g)2^{-1}}x_4^{2^h}.$$

If $x_4^{2^h}=1$, then x is the $(1+Y^*)$th power of some element in M, and $x=1$ is a consequence of the second part of condition (2).

If $h=1$, then $x=v^{1+Y^*}x_4^2=(vx_4)^2$ for v an element in M and vx_4 an element in Y^*; and it follows from condition (3') that $x=1$.

If $1<h$, then $x=v^{1+Y^*}x_4^{2(2^{h-1}-1)2}=w^{1+Y^*}$ for some element w in M (the existence of w is seen as in case B.2.1) and we infer again from the second part of condition (2) that $x=1$. Thus it follows that the cross-cut of P and $H(2)$ is 1; and this completes the proof of the theorem.

REMARK. From condition (2) one derives readily that $x^{p^{m(p)}}=1$ for every element x in G_{Cp} such that $x^{p^{m(p)}}$ is an element of order a power of p in G_C.

CHAPTER IV. UNIQUENESS OF THE C-CHARACTER GROUP

The group B of C-characters of the group G shall be termed *complete, if* 1 *is the only element in G mapped upon* 1 *by every C-character in B*; and the complete group B of C-characters of G may be called *strictly complete, if B contains the principal genus*. The C-character group of G (that is, the group of

all the C-characters of G) is obviously complete if, and only if, G is C-complete. Thus there arises the problem[21] of determining conditions assuring that the C-character group of G is the only complete (the only strictly complete) group of C-characters of G; and we devote this chapter to a solution of this problem.

IV.1. **Criteria depending on the nature of the homomorphism C only.** If G is C-complete, then G is by Theorem I.3.1 an extension of the abelian group G_C by the abelian group G/G_C; and the homomorphism C effects an isomorphism of G/G_C into the group of automorphisms of the cyclic group E. The criteria in this section will depend (like the investigations of Chapter II) only on the nature of the isomorphism C of the quotient group G/G_C.

THEOREM IV.1.1. *If the group G is C-complete, if the order of G_{C_p}/G_C is prime to p whenever $0<m(p)$ and G/G_{C_p} is of order a power of p, and if $C2$ is not singular, then the C-character group of G is the only strictly complete group of C-characters of G.*

Proof. If B is a complete group of C-characters of G, then it follows from the C-completeness of G and from Theorems I.3.1 and I.1.4 that every C_0-character of G_C in E is induced by some C-character in B. Thus there exists to every C-character f of G at least one C-character b in B such that f and b coincide on G_C. Since fb^{-1} is a C-character of G which maps G_C upon 1, it follows from our hypotheses and Theorem III.2.3 that fb^{-1} belongs to the principal genus. If B is strictly complete, then fb^{-1} is in B, showing that f itself belongs to B, that is, every strictly complete group B of C-characters of G is the (full) C-character group of G.

THEOREM IV.1.2. *If G is C-complete, if C is regular, then the C-character group of G is the only complete group of C-characters of G.*

Proof. It is an immediate consequence of Corollary II.5.3 and our hypothesis that G and its (full) C-character group are of equal order N. Suppose now that B is a complete group of C-characters of G. Then its order N^* certainly does not exceed N [is a divisor of N].

If g is an element in G, then we define a single-valued B to E function[22] $F_g(f)$ by the rule $F_g(f)=f(g)$ for every f in B. It is readily verified that F_g is a C_0-character of B in E; and we note that B is an abelian group the orders of whose elements are divisors of the order m of E. The elements x and y in G satisfy $F_x=F_y$ if, and only if, $f(x)=f(y)$ for every f in B; and this is equivalent to

$$1 = [f(x)f(y)^{-1}]^{\nu^{-1}} = f(x)^{\nu^{-1}}f(y)^{-\nu^{-1}} = f(x)^{\nu^{-1}}f(y^{-1}) = f(xy^{-1})$$

---

[21] A solution of this problem is important by its applications to the theory of radical extensions of fields, as will be shown elsewhere.

[22] Cp. V.1 below for a more detailed study of this operator F_g.

for every f in B. Since B is a complete group of C-characters of G, it follows that $F_x = F_y$ if, and only if, $xy^{-1}=1$ or $x=y$. Thus we have shown that the set F_G of all the C_0-characters F_g of B consists of exactly N characters. It is a consequence of Theorem I.1.3 that the order of the group of all the C_0-characters of B in E is exactly N^* showing that $N \leqq N^*$. But we have pointed out before that N^* does not exceed N showing that $N=N^*$, that is, B and the (full) C-character group of G are of equal order and are therefore equal, as was to be shown.

We note that the methods employed in the proofs of the two preceding theorems are completely exhausted by these theorems.

The following lemma will prove useful in the future.

LEMMA IV.1.3. *If G is C-complete, if G^* is a subgroup of G such that $G=G_C G^*$, if B is a group of C-characters of G such that every C-character of G^* is induced by some C-character in B, and if to every element w in G_C whose order modulo*([23]) *G^* is a prime there exists a C-character in B which maps G^* but not w upon 1, then B is the full C-character group of G.*

Proof. If f is a C-character of G, then there exists a C-character b in B such that f and b coincide on G^*. Thus $f_1=fb^{-1}$ is a C-character of G which maps G^* upon 1. If B^* is the group of all the C-characters in B which map G^* upon 1, then it follows from our hypotheses that G_C^* is the set of all the elements in G_C which are mapped upon 1 by all the characters in B^*. Hence it follows from Lemma I.3.4 and Theorem I.1.4 that every C_0-character of G_C in E which maps G_C^* upon 1 is induced by one and only one C-character in B^*; and thus it follows that the character f_1, constructed above, belongs to B^*. Hence $f=f_1 b$ belongs to B; and this completes the proof.

IV.2. **The case of primary E and cyclic G/G_C.** Throughout this section we impose the following hypotheses([24]):

(IV.2.*) E is a cyclic group of prime power order $m=p^n$, C is a homomorphism of the group G into the group of automorphisms of E, and G is C-complete.

These hypotheses imply the cyclicity of G/G_C, if m is odd, but not in the case of even m. Thus we add the following requirement:

(IV.2.**) G/G_C is cyclic.

THEOREM IV.2.1. *The C-character group of G is the only strictly complete group of C-characters of G if, and only if, C is not singular.*

Proof. If the condition of the theorem is satisfied, then the hypotheses of Theorem IV.1.1 are satisfied too, showing the sufficiency of our condition.

([23]) More precisely, modulo the cross-cut of G^* and G_C.

([24]) Without restating them in the hypotheses of the various theorems of this section.

If the condition is not satisfied, then $p=2$, $1<\mu$ and there exists in G an element z inducing the inversion in E. Because of the cyclicity of G/G_C it follows that z generates G modulo G_C; and $z^2=1$ is a consequence of the C-completeness of G, as follows from Theorem III.1.2. We denote by B the group of all those C-characters of G which map z upon an element in E^2. From $v^{1-z}=v^2$ for v in E we infer that B contains the principal genus. From Lemma I.3.4 and $z^2=1$ we deduce that every C_0-character of G_C in E is induced by one and only one C-character of G which maps z upon 1; and this implies that every C_0-character of G_C in E is induced by some C-character in the group B. Hence it follows from the C-completeness of G and Theorems I.3.1 and I.1.4 that there exists to every element $y\neq 1$ in G_C a C-character in B which does not map y upon 1; and if y is an element in G though not in G_C, then y is mapped upon elements different from 1 by the principal genus which is part of B. Thus B is a strictly complete group of C-characters of G. But we infer from Lemma I.3.4 and $z^2=1$ the existence of a C-character of G which maps G_C upon 1 and the element z upon an element not in E^2, that is, B is not the full C-character group of G as was to be shown.

THEOREM IV.2.2. *The C-character group of G is the only complete group of C-characters of G if, and only if, one of the following (mutually exclusive) conditions is satisfied:*

(1) *If p is odd and $G\neq G_C$, then either $m=p$ or the order of G/G_C is divisible by p.*

(2) *If $p=2$, G/G_C of order 2, and if C is irregular, then $G_C^2<G^2$.*

Proof. If $G=G_C$, then the C-characters of G are C_0-characters of G in E; and it follows from Theorem I.1.4 that the character group of G is the only complete group of characters. Consequently we assume in the following that G and G_C are different. This implies in particular that m is neither 1 nor 2.

If C is regular, then we infer from Theorem IV.1.2 that the C-character group of G is the only complete group of C-characters of G. Consequently we assume in the following that C is irregular.

If $p=2$, and if there exists in G an element z inducing the inversion in E, then G/G_C is of order 2, since it is cyclic; and $G^2=G_C^2$, since every element in G, not in G_C, induces the inversion in E, and since it follows from the C-completeness of G and Theorem III.1.2 that $y^2=1$, if y induces the inversion in E. On the other hand we infer from Theorem IV.2.1 that the C-character group of G is not the only complete group of C-characters of G. Consequently we may assume in the following that G does not contain an element inducing the inversion in E, if $p=2$.

Since G/G_C is cyclic, there exists an element g in G which generates G modulo G_C. If p is odd, then the automorphism induced by g in E is not of order a power of p; and we infer from II.1.A that the integer $g-1$ is prime to

p. The automorphism g of E does not leave invariant any element different from 1; and hence it follows from the C-completeness of G and from Theorem III.1.2 that 1 is the only element in G_C permuting with g. This implies in particular that the element g and the automorphism g of E are of equal order n.

If $p=2$, then the integer g corresponding to g under C is of the form $g\equiv -1+2^j g'$ modulo m for g' an odd number and $1<j<\mu$, since the integer corresponding to a square in G cannot be congruent to -1 modulo 4, and since G does not contain an inversion of E. We infer from II.1.B that the automorphism g of E is of order $2^{\mu-j}$. Since the automorphism g of E leaves only the elements in E whose order is a divisor of 2 invariant, it follows from the C-completeness of G and from Theorem III.1.2 that g permutes only with those elements in G_C whose orders are divisors of 2; and this implies in particular that the element $g^*=g^{2^{\mu-j}}$ in G_C is either 1 or of order 2.

We distinguish several cases.

Case 1. *If p is odd, then condition* (1) *is satisfied by G; and if $p=2$, then the order of G/G_C is divisible by* 4. We prove some lemmas.

(1.A) *Every complete group B of C-characters of G contains a character b such that $b(g^i)=1$ implies $g^i=1$.*

The complete group B of C-characters of G induces in the cyclic group $\{g\}$ generated by the element g a complete group $\overline{B}$ of C-characters. If p is odd and $m=p$, then we infer from Theorem II.5.2 that the C-character group of $\{g\}$ is a cyclic group of order p which is clearly generated by every element not 1. Since $\overline{B}$ cannot be 1, there exists in B a character b such that $b(g)\neq 1$ and this character b generates the full C-character group of $\{g\}$ and has therefore the desired qualities. Assume now that $m\neq p$. Then there exists a divisor n' of the order of the automorphism g of E such that $g''=g^{n'}$ is an automorphism of order p of E. It is a consequence of II.1 that the integer corresponding to g'' under C is of the form $g''\equiv 1+p^{\mu-1}g'''$ modulo m for g''' an integer prime to p. There exists in B a character b such that $b(g'')\neq 1$; and if $p=2$, but $g^*\neq 1$, then b may and shall be required to satisfy $b(g^*)\neq 1$. If $p=2$, $g^*=1$, then g'' is the only element of order 2 in $\{g\}$; and if $p=2$, $g^*\neq 1$, then g^* is the only element of order 2 in $\{g\}$; and thus $p=2$ and $b(g^i)=1$ imply that $g^i=1$. If p is odd, then $\{g\}_C=1$, as has been pointed out before; and it is a consequence of Theorem II.3.2 that every C-character of $\{g\}$ belongs to the principal genus. Hence there exists an element v in E such that $b(g^i)=v^{1-g^i}$. Since $1\neq b(g'')=v^{-p^{\mu-1}g'''}$, it follows that v is of order p^μ and that therefore v generates E. If $1=b(g^i)$ for some i, then g^i consequently leaves every element in E invariant; and this implies $g^i=1$, since the element g in G and the automorphism g of E have the same order n.

(1.B) *If the C-character d of* $\{g\}$ *does not map any element except* 1 *upon* 1, *then every C-character of* $\{g\}$ *is a power of d.*

If $m=p$, then the C-character group of $\{g\}$ is by Theorem II.5.2 a cyclic group of order p which is generated by each of its elements excepting 1. If $m\neq p$, and if $g^*=1$ in case $p=2$, then it follows from Theorem II.3.2 that every C-character of $\{g\}$ belongs to the principal genus, that therefore $d(g^i)=v^{1-s^i}$ for v in E, and that v is an element of order m in E, since $1\neq d(g'')$ for $g''=g^{n'}$ an automorphism of order p of E; and it is readily seen that every C-character of $\{g\}$ is a power of d. (Note that $g''\equiv 1+p^{n-1}g'''$ modulo m for g''' an integer prime to p.) If $p=2\neq m$, and if $g^*\neq 1$, then[25] one verifies in the same way that d^2 generates the principal genus of $\{g\}$ and that therefore every C-character of $\{g\}$ is a power of d, since it coincides with d or d^2 on g^*.

Suppose now that B is a complete group of C-characters of the group G. Then we denote by B_0 the group of all those C-characters in B which map g upon 1.

Consider an element w in G_C whose order modulo $\{g\}$ is p. If p is odd then the cross-cut of $\{g\}$ and G_C is 1 so that w is an element of order p in G_C. If $p=2$, then the cross-cut of $\{g\}$ and G_C is either 1 or generated by the element g^* of order 2; thus w is either of order 2 or $w^2=g^*$ and w is of order 4.

It is a consequence of (1.A) that there exists in B a C-character b such that $b(g^i)=1$ implies $g^i=1$; and we deduce from (1.B) that every C-character of $\{g\}$ is induced by some power of b.

If $b(w)=1$, then we deduce from the completeness of B the existence of a C-character f in B such that $f(w)\neq 1$; and there exists (by the remark made in the preceding paragraph) an integer k such that f and b^k coincide on $\{g\}$. Clearly $f_1=fb^{-k}$ belongs to B_0 and satisfies $f_1(w)=f(w)b(w)^{-k}=f(w)\neq 1$, showing that w is not mapped upon 1 by B_0.

If $b(w)\neq 1$, but w is of order p, then $b(w)$ is of order p, since b induces a C_0-character in G_C. There exists an integer r such that $b(g^r)$ is of order p; for if $m=p$ and p is odd, then take $r=1$; if p is odd, but $m\neq p$, then choose r in such a fashion that g^r induces an automorphism of order p in E, and a like choice is appropriate if $p=2$ and $g^*=1$; if $g^*\neq 1$, then $b(g^*)$ is of order 2. Since $b(g^r)$ and $b(w)$ are of equal order p, there exists an integer s, prime to p, such that $b(g^r)=b(w)^s=b(w^s)$; and we have $b(g^rw^{-s})=b(g^r)b(w)^{-s}=1$. From the completeness of B we infer the existence of a C-character f_2 in B such that $f_2(g^rw^{-s})\neq 1$; and there exists an integer k' such that f_2 and $b^{k'}$ coincide on $\{g\}$. Clearly $f_3=f_2b^{-k'}$ is a C-character in B_0, satisfying $f_3(w^{-s})=f_3(g^rw^{-s})$

---

[25] Another proof of this fact: from $d(g^*)=d(g'')^{1+g''}=d(g'')^2$ one derives that d is a C-character of order m. But it follows from Theorem II.5.2 that m is exactly the order of the C-character group of the cyclic group $\{g\}$.

$=f_3(g^r w^{-s})b(g^r w^{-s})^{-k'}=f_3(g^r w^{-s})\neq 1$; and from $f_3(w)^{-s}=f_3(w^{-s})\neq 1$ we infer $f_3(w)\neq 1$, showing that again w is not mapped upon 1 by B_0.

If $b(w)\neq 1$, but w is not of order p, then $p=2$, $g^*\neq 1$ and $w^2=g^*$. Denote by g'' an element in $\{g\}$ inducing in E an automorphism of order 2. Then $g^*=g''^2$, and the integer corresponding to g'' is of the form $g''\equiv 1+2^{\mu-1}$ modulo m. Since $b(g^*)=b(g''^2)=b(g'')^{1+\sigma''}=b(g'')^{2+2^{\mu-1}}$ is of order 2, it follows that $b(g'')$ is of order 4 whereas $b(w)$ is of order 2 or 4 and we find again an integer r such that $b(w)$ and $b(g^r)$ are of equal order, an odd integer s such that $b(w)^s=b(g^r)$; and we may now show, exactly as in the preceding paragraph, that w is not mapped upon 1 by B_0.

Thus we have shown that all the hypotheses of Lemma IV.1.3 are satisfied by the group B; and hence it follows that B is the full C-character group of G.

Case 2. $p=2$, the order of G/G_C is 2 and condition (2) *is satisfied by G.* Then the integer g corresponding to the element g in G is of the form $g\equiv -1+2^{\mu-1}$ modulo m; and $g^*=g^2$ is an element in G_C such that $g^{*2}=1$. Since $G^2\leqq G_C$, and since every element in G, not in G_C, has the form xg for x in G_C, it follows from $(xg)^2=x^{1+\sigma}g^*=x^{2^{\mu-1}}g^*$ that G^2 is obtained by adjoining g^* to G_C^2; and it follows from (2) that g^* is not a square in G_C so that in particular $g^*\neq 1$.

There exists in every complete group B of C-characters of G a C-character b such that $b(g^*)\neq 1$. Since g^* is an element of order 2 in G_C, $b(g^*)$ is an element of order 2; and we deduce from $b(g^*)=b(g^2)=b(g)^{1+\sigma}=b(g)^{2^{\mu-1}}$ that $b(g)$ is an element of order $2^\mu=m$ in E, that is $b(g)$ generates E; and this implies in particular that every C-character of $\{g\}$ is induced by a power of b.

If w is an element in G_C whose order modulo $\{g\}$ is 2, then $w^2=1$, since $w^2=g^*$ contradicts (2). Clearly $b(w)$ is either an element of order 2, as is $b(g^*)$, or is 1, as is $b(g^4)$; and thus there exists an integer r such that $b(g^r)$ and $b(w)$ are of equal order. Hence there exists an odd integer s such that $b(w)^s=b(g^r)$. Now we may show by the same arguments, as used in Case 1 above, that there exists a C-character in B which maps g upon 1 and w upon an element not 1; and we infer from Lemma IV.1.3 that the complete group B of C-characters of G is the full C-character group of G.

This completes the proof of the sufficiency of the conditions of our theorem.

Case 3. p is odd, but condition (1) *is not satisfied by G.* Then $1<\mu$ and the order of G/G_C is different from 1 and prime to p. It is a consequence of II.1 that the integers corresponding to elements y, not in G_C, meet the requirement $y-1$ is prime to p.

Consider the group B of all the C-characters f of G, satisfying $f(g)$ is in E^p.

Since the cross-cut of $\{g\}$ and G_C is 1, it follows from the C-completeness of G and Lemma I.3.4 that every C_0-character of G_C in E is induced by one and only one C-character in B which maps g upon 1. Hence it follows from

Theorem I.1.4 that every element y in G which is not in $\{g\}$ is mapped upon elements, not 1, by C-characters in B, since $y=xg^i$ for $x\neq 1$ in G_C. If v is an element, not 1, in E^p, then v^{1-e} is a C-character in B satisfying $v^{1-g^i}\neq 1$ for $g^i\neq 1$; and thus we have shown the completeness of the group B.

But B is not the full C-character group of G, since the C-character e^{1-e} for e a generator of E cannot be in B, as e^{1-g} is not in E^p. Thus we have shown that condition (1) is a necessary condition.

Case 4. $p=2$, *but condition* (2) *is not satisfied by* G. Then G/G_C is of order 2; the integer g corresponding to the element g in G satisfies $g\equiv -1+2^{\mu-1}$ modulo m; and the element $g^*=g^2$ whose order is 1 or 2 belongs to G_C^2; and $2<\mu$, since g is not in G_C.

Case 4.1. $g^*=1$. Consider the group B of all the C-characters of G which map g upon an element in E^4.

Then it is readily verified that there exists to every element not in $\{g\}$ a C-character in B which does not map it upon 1. If v generates E, then the C-character $v^{2(1-e)}$ is in B, since $v^{2(1-g)}=v^4$; and g is not mapped upon 1 by this C-character, since $v^4\neq 1$. But B does not contain the C-character v^{1-g}, that is, B is a complete group of C-characters which is not the full C-character group.

Case 4.2. $g^*\neq 1$. Then g^* is of order 2 and there exists an element w in G_C such that $w^2=g^*$. There exists, because of the C-completeness of G, a C-character f of G which maps g^* upon an element of order 2; and we denote by B the group of C-characters of G which is generated by f and by those C-characters of G which map g and w upon 1.

Since every C_0-character of G_C which maps w (and therefore g^*) upon 1 is induced by one and only one C-character in B which maps g upon 1, it follows that there exists to every element y in G which is not in the subgroup W generated by g and w a C-character in B which does not map y upon 1.

Since $f(g^*)$ is of order 2, and since $f(g^*)=f(g^2)=f(g)^{1+g}=f(g)^{2^{\mu-1}}$, it follows that $f(g)$ is of order m. Since the order 4 of w and $f(w)$ is smaller than m, and since every element in W, not in G_C, is of the form gw^i, it follows that $f(gw^i)=f(g)f(w)^i\neq 1$, that the elements, not in W_C, are mapped upon elements, not 1, by C-characters in B. Since W_C is generated by w, and since $f(w^2)=f(g^*)\neq 1$, we see finally that B is a complete group of C-characters of G.

From the argument used in the second paragraph of this Case 4.2 and from the fact that the order of the C-character f is m, since $f(g)$ is of order m, it follows that the order of B is exactly $2^{\mu-2}$ times the order of G_C. But it follows from Theorem II.5.2 that the order of the full C-character group of G is exactly $2^{\mu-1}$ times the order of G_C so that B is not the full C-character group of G.

This completes the proof of the necessity of condition (2); and we have completed the proof of our theorem.

IV.3. **The case of primary E and non-cyclic G/G_C.** Since a primary cyclic group whose group of automorphisms is not cyclic is of necessity of order a power of 2 [at least 8], we shall impose throughout this section the following hypotheses(28).

(IV.3.*) E is a cyclic group of order $m=2^\mu$ where $2<\mu$, C is a homomorphism of the group G into the group of automorphisms of E such that G/G_C is not cyclic, and G is C-complete.

It will be convenient to derive first some general properties of such groups and their C-characters.

LEMMA IV.3.1. (a) *There exists a basis g, z of G modulo G_C such that z induces an inversion in E and such that the integer corresponding to the element g under C is of the form $g\equiv -1+2^j g_0$ modulo m where g_0 is odd and $1<j<\mu$.*

(b) *If g, z is a basis of G modulo G_C, meeting the requirements of* (a), *if $n=2^\nu$ is the order of the automorphism g of E (and therefore half the order of G/G_C), if $g^*=g^n$ and G^* the group generated by g and z, then $z^2=g^{*2}=1$, $G=G_CG^*$, G_C^* is generated by g^* and the commutator $[g, z]$ whose order 2^h is a divisor of $2^{\mu-1}$.*

(c) *If y, z are elements in G such that z induces the inversion in E, and if f is a C-character of the group generated by y and z, then $f([z, y])^v=f(z)^{1-v}f(y)^{-2}$; and the C-character f^2 belongs to the principal genus of the group generated by y and z if, and only if, $f([z, y])=1$.*

It should be noted that the *commutator* of two elements v, w is defined by $[v, w]=vwv^{-1}w^{-1}$.

Proof. It is an immediate consequence of our hypotheses (IV.3.*) and of II.1.B that there exists a basis y, z of G modulo G_C such that z is an inversion of E; and it is readily verified that either the basis y, z or the basis yz, z of G modulo G_C meets the requirements of (a).

If the basis g, z of G modulo G_C meets the requirements enunciated in (a), then the automorphism g of E leaves only the elements of an order dividing 2 invariant; and it follows from the C-completeness of G and Theorem III.1.2 that $z^2=g^{*2}=1$. Since the group of automorphisms of E which are induced by elements in G is the direct product of a cyclic group of order 2 by a cyclic group of order n, and since this group is isomorphic to G^*/G_C^*, it follows that G_C^* is generated by g^* and the commutator $[g, z]$. This completes the proof of (b) with the exception of the assertion concerning the order of $[g, z]$.

Suppose now that the element z in G induces the inversion in E, that y is some element in G, that Y is the subgroup generated by y and z, and that f is some C-character of Y. Then

(28) Without restating them in the hypotheses of the various theorems of this section.

$$f(z)^{y-1}f(y)^2 = f(z)^y f(y)\,[f(y)^{-1}f(z)]^{-1} = f(zy)f(yz)^{-1}$$
$$= f(zy)f[(yz)^{-1}]^y = f[(yz)^{-1}zy] = f(zy^{-1}zy);$$

and substituting y^{-1} for y we obtain:

$$f(zyzy^{-1}) = f(z)^{y^{-1}-1}f(y^{-1})^2 = f(z)^{y^{-1}-1}f(y)^{-2y}$$

or

$$f([z, y])^y = f(z)^{1-y}f(y)^{-2}.$$

Since y is odd and $1-y$ is even, it follows from this formula that $f([z, y]^{2^{s-1}}) = f([z, y])^{2^{s-1}} = 1$; and we infer from the C-completeness of G that $[z, y]^{2^{s-1}} = 1$, completing the proof of (b).

If $f^2(x) = v^{1-s}$ for v in E and x in Y, then we infer from the above formula and the evenness of $1-y$ that $f([z, y])^y = v^{(1-s)(1-y)2^{-1}-(1-y)} = 1$ so that $f([z, y]) = 1$; and if conversely $f([z, y]) = 1$, then it follows from the above formula that $f(y)^2 = f(z)^{1-y}$; and it follows from Theorem II.3.1(2) that f^2 belongs to the principal genus of Y.

LEMMA IV.3.2. *If g, z is a basis of G modulo G_C, and if z induces the inversion in E, then the commutator subgroup G' of G is obtained by adjoining the commutator $[g, z]$ to G_C^2.*

Proof. If x is in G_C and y in G, then the commutator $x^{1-y} = [x, y^{-1}]$ is in G_C^2, since $1-y$ is even; and in particular $x^{1-z} = x^2$. Now it is clear how to complete the proof, since subgroups of G_C are by Theorem I.3.1 and Corollary I.3.2 normal subgroups of G.

LEMMA IV.3.3. *If $G' = G_C^2$, then there exists a strictly complete group of C-characters of G which is not the full C-character group of G.*

Proof. Suppose that the elements g, z in G form a basis of G modulo G_C which meets the requirements of Lemma IV.3.1(a). It is a consequence of our hypothesis that the commutator $[g, z]$ belongs to G_C^2 and that there exists therefore an element w in G_C, satisfying $(w^{g^{-1}(1-2^{f-1}g_0)})^2 = [z, g]$, since $g^{-1}(1-2^{f-2}g_0)$ is odd. It is clear that the elements g, wz form a basis of G modulo G_C which meets the requirements of Lemma IV.3.1(a); and one finds that

$$[g, wz] = gwzg^{-1}(wz)^{-1} = w^{g^{-1}}[g, z]w^{-1} = w^{g^{-1}-1}[g, z] = 1.$$

Hence we may assume without loss of generality that the basis g, z of G modulo G_C satisfies beyond the conditions of Lemma IV.3.1(a) the additional condition $[g, z] = 1$ so that the subgroup G^* of G which is generated by g and z is *an abelian subgroup*. It is a consequence of Lemma IV.3.1(b)

that G_C^* is generated by the element $g^* = g^n$ for n the order of the automorphism g of E, and that $g^{*2} = 1$.

Case 1. $g^* = 1$. Then we consider the group B of C-characters of G which is generated by the principal genus of G and by all those C-characters of G which map G^* upon 1. Since B contains the principal genus, elements not in G_C are not mapped upon 1 by all the C-characters in B. Since 1 is the cross-cut of G^* and G_C, it follows from Lemma I.3.4 that every C_0-character of G_C in E is induced by one and only one C-character of G which maps G^* upon 1 and which therefore belongs to B; and thus it follows from the C-completeness of G and Theorem I.1.4 that 1 is the only element in G_C (and therefore in G) which is mapped upon 1 by all the C-characters in B. Thus B is shown to be strictly complete. Since B is the direct product of the principal genus and of the group of C-characters of G which map G^* upon 1, and since the order of the latter group is by Theorem I.1.3 and the C-completeness of G (and by a previous remark) exactly the order of G_C, it follows from Theorem II.2.2(a) that the order of B is $2^{\mu-1}$ times the order of G_C. On the other hand we infer from Theorem II.5.2 that the order of the full C-character group of G is exactly 2^μ times the order of G_C showing that B cannot be the full C-character group of G.

Case 2. $g^* \neq 1$. The cross-cut G_C^* of G^* and G_C is in this case by Lemma IV.3.1(b) the cyclic group of order 2 which is generated by g^*.

Denote by B^* the group of all C-characters of G which map G^* upon 1. It is an immediate consequence of the C-completeness of G, of Lemma I.3.4 and Theorem I.1.4 that every C_0-character of G_C in E which maps g^* upon 1 is induced by one and only one C-character in B^*; and it follows from Theorem I.1.3 that the order of B^* is half the order of G_C; and that the elements not in G^* are mapped upon elements not 1 by characters in B^*.

Denote by B^{**} the group of C-characters of G which is generated by adjoining the principal genus of G to B^*. The cross-cut of B^* and the principal genus of G is evidently 1. Thus the order of B^{**} is $2^{\mu-2}$ times the order of G_C, since the order of the principal genus is by Theorem II.2.2(a) exactly $2^{\mu-1}$. Clearly the elements, not in G_C^*, are mapped upon elements different from 1 by suitable C-characters in B^{**}.

Since G is C-complete, there exists a C-character f of G such that $f(g^*) \neq 1$. This C-character f does not belong to B^{**}, since the characters in B^* as well as in the principal genus map g^* upon 1. From $[g, x] = 1$ we infer $f([g, x]) = 1$; and hence it follows from Lemma IV.3.1(c) that f^2 induces in G^* a character of the principal genus of G^*. There exists therefore an element v in E such that $f^2(x) = v^{1-x}$ for x in G^* so that the C-character $b(x) = f^2(x)v^{x-1}$ belongs to B^*. Hence f^2 itself belongs to B^{**}; that is, f is of order 2 modulo B^{**}. We denote by B the group of C-characters of G which is generated by adjoining f to B^{**}. It follows from what we have shown thus far that B is a strictly complete

group of C-characters of G; and that the order of B is $2^{\mu-1}$ times the order of G_C, since the order of B is twice the order of B^{**}. But we infer again from Theorem II.5.2(a) that the order of the C-character group of G is exactly 2^μ times the order of G_C showing that the strictly complete group B of C-characters of G is not the full C-character group of G; and this completes the proof.

It is a consequence of Lemma IV.3.1(a) that G/G_C is the direct product of a cyclic group Z and a cyclic group Z' where the elements in Z induce in E the inversion (so that Z is of order 2) whereas Z' is generated by an element to which corresponds under C an integer of the form $-1+2^j g_0$ modulo m where g_0 is odd and $1<j<\mu$. It is readily seen that Z and Z' are uniquely determined by these properties. The normal *subgroup* $A=A(G, C)$ of G which contains G_C and satisfies $Z'=A/G_C$ is consequently uniquely determined by G and C.

LEMMA IV.3.4. *If the basis g, z of G modulo G_C meets the requirements of Lemma* IV.3.1(a), *if n is the order of the automorphism g of E and $g^*=g^n$, then A is obtained by adjoining g to G_C; and the cross-cut of A^2 and G_C is generated by g^* and G_C^2.*

Proof. It is clear that the coset $G_C g$ is a generator of the cyclic group $Z'=A/G_C$; and that therefore A is obtained by adjoining g to G_C. The elements in A whose squares are in G_C are clearly contained in the cosets G_C and $G_C g^{n^{2^{-1}}}$ of A/G_C, as these make up the cyclic subgroup of order 2 of Z'. The cross-cut of the groups A^2 and G_C therefore certainly contains the elements in G_C^2 and the element $(g^{n^{2^{-1}}})^2=g^*$. Any further element is of the form $(xg^{n^{2^{-1}}})^2=x^{1+\sigma^{-n^{2^{-1}}}}g^*$ for x in G_C; and this element is contained in the group generated by adjoining g^* to G_C^2, since the exponent of x is certainly an even integer; and this completes the proof.

LEMMA IV.3.5. *If the order of G/G_C is not* 4 *(and therefore divisible by* 8), *if the commutator subgroup $G'\leqq A^2$, then there exists a complete group of C-characters of G which is not the full C-character group of G.*

Proof. This is an immediate consequence of Lemma IV.3.3, in case $G'=G_C^2$; and hence we may assume that $G_C^2<G'\leqq A^2$. Since $G'\leqq G_C$, and since the order of G'/G_C^2 as well as the order of[27] $(G_C\cap A^2)/G_C^2$ is at most 2 (as follows from Lemmas IV.3.2 and IV.3.4), it follows that $G_C^2<G'=G_C\cap A^2$. If g, z is some basis of G modulo G_C which meets the requirements of Lemma IV.3.1(a), then it follows from Lemmas IV.3.2 and IV.3.4 that $[g, z]\equiv g^*$ modulo G_C^2 where $g^*=g^n$ for n the order of the automorphism g of E; and that neither $[g, z]$ nor g^* is in G_C^2 (so that g^* is, by Lemma IV.3.1(b), of order 2). Since the integer $(1-g)g^{-1}$ is of the form $(1-g)g^{-1}\equiv 2(1-2^{j-1}g_0)g^{-1}$

[27] $S\cap T$ denotes the cross-cut of the sets S and T.

modulo m where the factor of 2 is odd, there exists an element w in G_C such that $g^* = w^{(1-g)g^{-1}}[g, z] = [g, wz]$. Since the elements g, wz form, exactly as g, z, a basis of G/G_C which meets the requirements of Lemma IV.3.1(a), we may assume without loss of generality that g, z is a basis of G modulo G_C which meets the requirements of Lemma IV.3.1(a) and which satisfies $[g, z] = g^*$.

We note that the order of the automorphism g of E is a multiple of 4 (or $1 < r$), since the order of G/G_C is a multiple of 8.

Since g^* is an element of order 2 in G_C which is not a square in G_C there exists a C_0-character f_1 of G_C in E which maps G_C^2, but not g^* upon 1 (cp. Theorem I.1.4!); and it is clear that $f_1^2 = 1$. Since G is C-complete, we infer from Theorem I.3.1 the existence of a C-character f of G which induces f_1 in G_C. Since n is the order of the automorphism g of E, and since g^* and $f(g^*)$ are elements of order 2, we infer from II.1.B that

$$f(g^*) = f(g^n) = f(g)^{1+g+\cdots+g^{n-1}} = f(g)^{2^{\mu}-1};$$

and this shows that $f(g)$ is an element of order 2^{μ}, so that $f(g)$ generates E. From Lemma IV.3.1(c) we infer that

$$f(g^*)f(g)^2 = f([z, g])^g f(g)^2 = f(z)^{1-g} = f(z)^{2-2^{\mu_0}}$$

showing that $f(z)^2$ is of order $2^{\mu-1}$ and $f(z)$ of order 2^{μ}.

The element $g^{n2^{-1}}$ induces in E an automorphism of order 2 and the integer corresponding to $g^{n2^{-1}}$ under C is of the form $1+2^{\mu-1}$ modulo m, since $n2^{-1}$ is even. Hence we find that

$$f(g^*) = f(g^{n2^{-1}z}) = f(g^{n2^{-1}})^{1+g^{n2^{-1}}} = f(g^{n2^{-1}})^{2+2^{\mu-1}} = f(g^{n2^{-1}})^2,$$

since $f(g^*)$ is of order 2 and since $1+2^{\mu-2}$ is odd.

The C-character f^2 of G maps G_C upon 1. If w is an element in G, not in G_C, whose order modulo G_C is two, then

$$w \equiv z \quad \text{or} \quad g^{n2^{-1}} \quad \text{or} \quad g^{n2^{-1}}z \text{ modulo } G_C.$$

Accordingly we find that $f(w)^2 = f(z)^2$ or $f(g^{n2^{-1}})^2$ or $f(g^{n2^{-1}})^{-2}f(z)^2$. In the first of these cases $f(w)^2$ is of order $2^{\mu-1}$; in the second case $f(w)^2$ is of order 2; and in the third case $f(w)^2$ is of order $2^{\mu-1}$, since it is the product of an element of order 2 by an element of order $2^{\mu-1}$. In all three cases $f(w)^2 \neq 1$; and this shows that f^2 maps an element upon 1 if, and only if, this element is in G_C.

Consider now the group B_1 of all the C-characters of G which map g and z upon 1. If G^* is the subgroup, generated by g and z, then the cross-cut G_C^* of G^* and G_C is, by Lemma IV.3.1(b), the cyclic subgroup of order 2 generated by g^*. Hence it follows from the C-completeness of G and Lemma I.3.4 that every C_0-character of G_C in E which maps g^* upon 1 is induced by one

and only one C-character in B_1. Thus the order of B_1 is half the order of G_C; and there exists to every element in G_C, not in G_C^*, a C-character in B_1 which does not map it upon 1.

Denote by B the group of C-characters of G, generated by f and B_1. The order of B is 2^μ times the order of B_1, since the order of $f(z)$ is 2^μ, that is, the order of B is $2^{\mu-1}$ times the order of G_C. The group B is complete, since elements not in G_C are mapped upon elements, not 1, by f, since the elements in G_C, not in G_C^*, are mapped by C-characters in B_1 upon elements different from 1, and since $f(g^*) \neq 1$. But it follows from Theorem II.5.2 that the order of the full C-character group of G is 2^μ times the order of G_C, showing that the complete group B of C-characters of G is not the full C-character group of G.

LEMMA IV.3.6. *If G/G_C is of order 4, if $A^2 = G_C^2$, then there exists a complete group of C-characters of G which is not the full C-character group of G.*

Note that $G^2 \leqq G_C$, since G/G_C is the direct product of two cyclic groups of order 2.

Proof. It is a consequence of Corollary III.1.3 that every C-character of the subgroup A of G is induced by some C-characters of G, since G is C-complete. It is a consequence of $A^2 = G_C^2$ and of Theorem IV.2.2(a) that there exists a complete group D of C-characters of A which is not the group of all the C-characters of A. Denote by B the group of all the C-characters of G which induce in A C-characters in D. Since not every C-character of A is induced by C-characters in B, and since every C-character of A is induced by some C-character of G, it follows that B is not the full C-character group of G.

There exists a C-character f of G which maps A upon 1 and the elements, not in A, upon the element of order 2 in E, since G/A is of order 2, and since the automorphisms of E leave the element of order 2 invariant. Clearly f belongs to B; and thus it follows that B is complete, since B induces in A the complete group D.

THEOREM IV.3.7. *The C-character group of G is the only complete group of C-characters of G if, and only if,*

(i) *G' is not part of A^2 in case the order of G/G_C is a multiple of 8; and*

(ii) *neither G' nor A^2 is equal to G_C^2 in case G/G_C is of order 4.*

Proof. The necessity of condition (i) is a restatement of Lemma IV.3.5 whereas the necessity of (ii) may be derived from Lemmas IV.3.3 and IV.3.6.

In the proof of the sufficiency of the conditions (i), (ii) we distinguish two cases:

A. $G' > A^2$. We use a basis g, z of G modulo G_C which meets the requirements of Lemma IV.3.1(a), denote by n the order of the automorphism g and put $g^* = g^n$. By G^* we denote the subgroup generated by g and z and by A^* the subgroup generated by g and $[g, z]$. Then it follows from Lemma

IV.3.1 that $G_C^* = A_C^*$ is generated by g^* and $[g, z]$ where $g^{*2} = 1$ and where the order of the commutator $[g, z]$ shall be denoted by 2^λ. It is a consequence of Lemmas IV.3.2 and IV.3.4 that the hypothesis A is equivalent to the two statements:

A'. $[g, z] \neq g^*$ modulo G_C^2 and $[g, z] \neq 1$ modulo G_C^2.

Suppose now that B be a complete group of C-characters of G. Then B induces both in A and in A^* a complete group of C-characters; and it is a consequence of condition (ii) and Theorem IV.2.2(2) that:

(*) *Every C-character of A, A^* and of the cyclic group $\{g\}$ generated by g is induced by some C-character in B.*

It is a consequence of Theorem II.5.2 that the order of the [cyclic] C-character group of[28] $\{g\}$ is 2^μ, if $g^* \neq 1$, and $2^{\mu-1}$, if $g^* = 1$. Hence there exists a C-character b_1 in B such that $b_1(g)$ is of order 2^μ, if $g^* \neq 1$, and of order $2^{\mu-1}$, if $g^* = 1$. If f is any C-character of G, then there exists an integer i such that $b_1(g)^i = f(g)$. Consequently there exists a C-character b_2 in B such that $b_2(g) = 1$ and such that the order of $b_2([g, z])$ is 2^λ, if the cross-cut of the cyclic groups $\{g\}$ and $\{[g, z]\}$ is 1, and such that the order of $b_2([g, z])$ is $2^{\lambda-1}$, if $g^* \neq 1$ is a power of $[g, z]$. It is a consequence of A' that $1 < h$, if $g^* \neq 1$ is a power of $[g, z]$ so that b_2 is different from 1 on A^*. These two C-characters b_1 and b_2 are independent on A^*; and the order of the group of C-characters of A^* induced by b_1, b_2 and their products is exactly $2^{\mu-1}$ times the order of A_C^* so that every C-character of A^* is induced by a combination of b_1 and b_2.

It is a consequence of Lemma IV.3.1(c) that

$$b_2([z, g]) = b_2(z)^{(1-g)g^{-1}} = b_2(z)^{2(1-2^{\lambda-1}g_0)g^{-1}}$$

and that therefore the order of $b_2(z)$ is twice the order of $b_2([z, g])$. The order of the C-character of G^*, induced by b_2, is therefore twice the order of the C-character of A^*, induced by b_2, since the latter is the same as the order of $b_2([z, g])$. The group of C-characters of G^*, induced by the combinations of b_1 and b_2, has therefore the order 2^μ times the order of G_C^* (note that $G_C^* = A_C^*$); and thus it follows from Theorem II.5.2 that

(**) *every C-character of G^* is induced by some combination of the C-characters b_1 and b_2 in B.*

Suppose now that W_0 is a coset of order 2 in G_C/G_C^* and that W is the subgroup of G, generated by adjoining W_0 to A^*. It is a consequence of hypothesis A' that W_0 contains an element w satisfying $w^2 = 1$ or $w^2 = g^*$ (the latter can happen only if the order n of the automorphism g of E is not 2). Since $b_2(g^*) = 1$, it follows that $b_2(w)$ is an element of order 1 or 2 showing that b_2 induces C-characters of equal order in W and in A^*. The order of

[28] It should be remembered that $\{g\}$ denotes the cyclic group, generated by the element g.

$b_1(w)$ is a divisor of 4 and 4 is a divisor of $2^{\mu-1}$ and therefore of the order of b_1, showing that b_1 induces C-characters of equal order in W and in A^*. Thus it follows that the group D of C-characters, generated by b_1 and b_2, induces in W a group of C-characters whose order is $2^{\mu-1}$ times the order of A_C^*. Since w is not in A_C^*, and is of order 2 modulo A_C^*, it follows from Theorem II.5.2 that the C-character group of W is of order 2^μ times the order of A_C^*; and hence there exists a C-character r of W which is not induced by C-characters in D. But $W \leqq A$; and hence it follows from (ii) and Theorem IV.2.2(2) that every C-character of W is induced by some C-character in B. In particular there exists a C-character b in B which induces r in W. We infer from (**) the existence of integers i', i'' such that b and $b_1^{i'}b_2^{i''}$ coincide on G^*; and we put $b_3 = bb_1^{-i'}b_2^{-i''}$. This C-character b_3 certainly belongs to the group B^* of all the C-characters in B which map G^* upon 1. Since b_3 cannot map every element in W upon 1, and since b_3 maps every element in A^* upon 1 (as A^* is part of G^*), it follows that $b_3(w) \neq 1$; and this implies $b_3(W_0) \neq 1$. Hence there exists to every element in G_C whose order modulo G^* (or $G_C^* = A_C^*$) is 2 a C-character in B^* which does not map it upon 1. Now we infer from (**) and Lemma IV.1.3 that B is the full C-character group of G, as was to be shown.

B. $G' \leqq A^2$. In this case we infer from (i) that the order of G/G_C is 4. Using condition (ii) we infer now—exactly as in the beginning of the proof of Lemma IV.3.5—the existence of a basis g, z of G modulo G_C meeting the requirements of Lemma IV.3.1(a) and satisfying:

B'. $g^2 = g^* = [g, z]$ is an element of order 2, not in G_C^2. We note that the integer corresponding under C to the element g in G is of the form $g \equiv -1 + 2^{\mu-1}$ modulo m; and we verify that

$$[g^{-1}, z] = g^{-1}zgz = g^*gzg^{-1}g^*z = g^*[g, z]g^* = [g, z].$$

Suppse now that B is a complete group of C-characters of G. Then there exists in B a C-character such that $b(g^*) \neq 1$. Hence $b(g^*) = b(g^2) = b(g)^{1+\sigma} = b(g)^{2^{\mu-1}}$ is an element of order 2; and $b(g)$ is an element of order 2^μ. From Lemma IV.3.1(c) we derive $b(g^*)b(g)^2 = b(g^*)^\sigma b(g)^2 = b(z)^{1-\sigma} = b(z)^{2+2^{\mu-1}}$. Since $b(g)^2$ is of order $2^{\mu-1}$ and $b(g^*)$ of order 2, we deduce from this last equation that $b(z)$ is of order 2^μ so that $b(z)^{2^{\mu-1}}$ is the only element of order 2 in E, namely $b(g^*)$; and hence it follows that $b(g)^2 = b(z)^2$.

If $b(g) = b(z)$, then $b(gz) = b(g)^{-1}b(z) = 1$. If $b(g) \neq b(z)$, then it follows from $b(g)^2 = b(z)^2$ that $b(z) = b(g)b(g^*) = b(gg^*) = b(g^{-1})$ since $b(g^*)$ is the element of order 2 in E; and this shows that $b(g^{-1}z) = b(g^{-1})^{-1}b(z) = 1$. In both cases there exists therefore in the group G^*, generated by g and z, an element which is not contained in G_C and which is mapped upon 1 by the C-character b. Since B induces a complete group of C-characters in G^*, there exists a C-character b' in B which does not coincide with any power of b on G^*. Denote by D the

subgroup of B, generated by b and b'. Since b is of order 2^μ on G^*, it follows that D induces in G^* a group of C-characters whose order is at least $2^{\mu+1}$. But this is by Theorem II.5.2 exactly the order of the C-character group of G^*, since G_C^* is by Lemma IV.3.1(2) and condition B′ the cyclic group of order 2, generated by g^*. Thus we have shown:

(+) *Every C-character of G^* is induced by some C-character in $D \leqq B$.*

Since $b(g)$ is an element of order 2^μ, there exists an integer i such that $b(g)^i = b'(g)$. Put $b_0 = b'b^{-i}$. Then b, b_0 generate D and b_0 is different from 1 on G^*, though $b_0(g) = 1$. Since this implies $b_0([\xi, z]) = b_0(g^*) = b_0(g^2) = 1$, it follows from Lemma IV.3.1(c) that $b_0(z)$ is an element of order 2.

We denote again by B^* the group of all those C-characters in B which map G^* upon 1.

If W_0 is a coset of G_C/G_C^* whose order is 2, then it follows from (ii) and condition B′ that W_0 contains an element w of order 2 which is not in G_C^*, that is $w \neq g^*$. The complete group B of C-characters of G induces in G_C a complete group of C_0-characters; and it follows from Theorem I.1.4 that every C_0-character of G_C in E is induced by some C-character in B. Hence it follows from Lemma I.1.1(b) that B contains a C-character which maps w, but not g^*, upon 1; and we may assume without loss of generality that the C-character b, considered before, has this property.

If $b_0(w) = 1$, then let f be any C-character in B such that $f(w) \neq 1$ (the existence of f is a consequence of $w \neq 1$ and of the completeness of B). There exist integers i, i_0 such that f and $b^i b_0^{i_0}$ coincide on G^*, as follows from (+); and hence $f^* = fb^{-i}b_0^{-i_0}$ is in B^*. But $f^*(w) = f(w) \neq 1$ so that B^* contains a C-character which does not map W_0 upon 1.

Suppose next that $b_0(w) \neq 1$. Then $b_0(w)$ is of order 2, that is $b_0(w) = b_0(z)$. It has been shown before that either $b(gz)$ or $b(g^{-1}z)$ is equal to 1. Hence determine $e = \pm 1$ in such a fashion that $b(g^e z) = 1$. Then $b(wg^e z) = 1$ and $b_0(wg^e z) = b_0(w)b_0(g^e z) = b_0(w)b_0(z) = 1$, since $b_0(g) = 1$, and since $b_0(w) = b_0(z)$ is of order 2. From the completeness of B we infer the existence of a C-character t in B such that $t(wg^e z) \neq 1$; and from (+) we deduce the existence of integers k, k_0 such that $t^* = tb^{-k}b_0^{-k_0}$ is in B^* (maps G^* upon 1). Hence $t^*(W_0) = t^*(w) = t^*(w)^{e^2}t^*(g^e z)$, since $t^*(w)$ is of order 2 and $t^*(g^e z)$ is equal to 1; and therefore we find that $t^*(W_0) = t^*(wg^e z) = t(wg^e z) \neq 1$, since $b(wg^e z) = b_0(wg^e z) = 1$; and we have shown again the existence of a C-character in B^* which does not map W_0 upon 1.

But now it is an immediate inference from (+) and Lemma IV.1.3 that B is the full C-character group of G; and this completes the proof of the theorem.

THEOREM IV.3.8. *The C-character group of G is the only strictly complete group of C-characters of G if, and only if, $G_C^2 < G'$.*

Proof. It is a consequence of Lemma IV.3.2 that $G_C^2 \leqq G'$. Hence we may deduce the necessity of our condition from Lemma IV.3.3.

Assume now that $G_C^2 < G'$. We distinguish two cases:

Case 1. *The order of* G/G_C *is* 4. If $A^2 \neq G_C^2$, then our contention is a consequence of Theorem IV.3.7. Thus we assume now that $A^2 = G_C^2$. Consider a basis g, z of G modulo G_C meeting the requirements of Lemma IV.3.1(a). Then $g^* = g^2$ is in G_C^2.

If B is a strictly complete group of C-characters of G, then B induces in A a strictly complete group of C-characters; and it follows from Theorem IV.2.1 that B induces in A every C-character of A. It is a consequence of the C-completeness of G and A and of Lemma I.3.4 that there exists a C-character of A which maps g and G_C^2 upon 1 but which maps $[g, z]$ upon an element of order 2, since g^* is in G_C^2 whereas $[g, z]$ does not belong to G_C^2; and hence there exists in B a C-character b which maps both g and G_C^2 upon 1 though $b([g, z])$ is of order 2. Hence it follows from Lemma IV.3.1(c) that $b(z)^{1-s} = b([z, g])^s$; and this implies $b(z)^2 = b([z, g])$, since $b([z, g])$ is an element of order 2. The C-character b^2 of G has therefore the properties $b^2(A) = 1$, $b^2(z)$ is of order 2. Hence it follows that every C-character of A is induced by at least two C-characters in B. The order of the C-character group of A is by Theorem II.5.2 exactly $2^{\mu-1}$ times the order of G_C; and the order of the C-character group of G is exactly 2^μ times the order of G_C; and from these facts one deduces that B is the full C-character group of G.

Case 2. *The order of* G/G_C *is divisible by* 8. If G' is not part of A^2, then our contention is a consequence of Theorem IV.3.7. Thus we assume now that $G_C^2 < G' \leqq A^2$; and we prove—exactly as in the beginning of the proof of Lemma IV.3.5—the existence of a basis g, z of G modulo G_C which meets the requirements of Lemma IV.3.1(a) and which satisfies $g^n = g^* = [g, z]$ is an element of order 2 (n denotes, as always, the order of the automorphism g of E).

If B is a strictly complete group of C-characters of G, then B induces in A a strictly complete group of C-characters; and it follows from Theorem IV.2.1 that every C-character of A is induced by some C-character in B.

There exists a C_0-character of G_C in E which maps G_C^2 upon 1 and g^* upon an element of order 2, since $g^* = [g, z]$ is not in G_C^2; and it follows from the C-completeness of G and A and from Theorem I.3.1 that this C_0-character of G_C is induced by some C-character of A. Hence there exists a C-character b in B such that $b(G_C^2) = 1$ and $b(g^*)$ is the element of order 2 in E. From $b(g^*) = b(g^n) = b(g)^{1+s+\cdots+s^{n-1}}$ and from II.1.B we infer $b(g^*) = b(g)^{2^{\mu-1}}$ is an element of order 2 so that $b(g)$ is an element of order 2^μ. Clearly $b^2(G_C) = 1$. But from $b([g, z]) \neq 1$ and Lemma IV.3.1(c) we infer that b^2 does not belong to the principal genus.

From our choice of g it follows that the integer corresponding to g under C is of the form $g \equiv -1 + 2^j g_0$ modulo m for g_0 an odd integer and $1 < j < \mu$.

The integer $1-g$ is therefore divisible by 2, but not by 4; and hence there exists an element v in E such that $b(g)^2=v^{1-g}$. The C-character v^{1-x} of G belongs to B; and B contains therefore the C-character $d(x)=b(x)^2v^{x-1}$. This C-character d is not 1, since b^2 has been shown not to be in the principal genus; but d maps both g and G_C upon 1; that is, the C-character $d\neq 1$ in B maps A upon 1. Now it follows that every C-character of A is induced by at least two C-characters in B. We deduce from Theorem II.5.2 that the order of the C-character group of G is exactly twice the order of the C-character group of A, showing that B is the full C-character group of G; and this completes the proof.

IV.4. **Completeness and p-completeness of groups of C-characters.** We have treated the problem of complete groups of C-characters completely for the special case of primary E. The reduction of the general case where E is of composite order to this special case will now be accomplished by means of the following considerations.

The group B of C-characters of the group G is the direct product of its subgroups B_p which consist of all the elements of order a power of (the prime number) p, since B is abelian. A C-character of G is of order a power of p if, and only if, it maps G into $E(p)$ (using the notations (II.3.*)!); and thus B_p consists exactly of all the Cp-characters in B. It is a consequence of the definition of the subgroup $H(p)$ that $B_p(H(p))=1$; and this leads to the following definition:

The group B of C-characters of G is p-complete, if $B_p(x)=1$ implies that x is in $H(p)$.

If the group B of C-characters of G is p-complete for every p, and if x is an element in G which is mapped upon 1 by B, then $B_p(x)=1$ for every p; and consequently x belongs to the cross-cut H of all the $H(p)$. If G is C-complete, and if B is p-complete for every p, then we deduce the completeness of B. The converse of this result need not be true; and it will be the main object of this section to find criteria for the validity of the converse.

THEOREM IV.4.1. *If G is C-complete, and if p is a prime, then the following condition is necessary and sufficient for the p-completeness of every strictly complate group of C-characters of G:*

(p) *$H(p)G_C$ contains every element[29] in G_{Cp} whose order modulo $H(p)$ is exactly p.*

Proof. Assume first that condition (p) is not satisfied. Then there exists an element w in G_{Cp} which is not contained in $H(p)G_C$, though its order modulo $H(p)$ is p. We denote by B the group of C-characters of G which is

[29] See Lemma III.2.2 and Corollary III.3.2 for properties of the subgroup $H(p)G_C=H^*(p)$.

generated by all the Cq-characters for $q \neq p$, and by all those Cp-characters which map w upon 1. This group B contains the principal genus, since w is in G_{Cp}. Thus $B(x)=1$ implies that x is in G_C. If $x \neq 1$ is an element of order a power of q for q a prime different from p, and if x is in G_C, then it follows from the C-completeness of G and from Theorem III.3.1 that x does not belong to $H(q)$. It is a consequence of the definition of $H(q)$ that $G/H(q)$ is Cq-complete; and hence there exists a Cq-character of G which does not map x upon 1. But all the Cq-characters are in B, that is $B(x) \neq 1$. If $x \neq 1$ is an element of order a power of p in G_C, then it follows again from the C-completeness of G and from Theorem III.3.1 that x is not in $H(p)$; and it follows from the fact that w is of order p modulo $H(p)$, but does not belong to $H(p)G_C$, that $H(p)w$ and $H(p)x$ are independent elements of $G_{Cp}/H(p)$. Hence we deduce from Lemma I.1.1(b) the existence of a C_0-character d of $G_{Cp}/H(p)$ in $E(p)$ which maps $H(p)w$ upon 1, but which does not map $H(p)x$ upon 1. It follows from the Cp-completeness of $G/H(p)$ and from Theorem I.3.1 that d is induced by a Cp-character of $G/H(p)$; and d is therefore induced by a Cp-character of G. This Cp-character of G belongs to B, since it maps w upon 1; and thus we have shown that $B(x) \neq 1$. From the facts derived thus far it follows that B is a strictly complete group of C-characters of G. But it follows from the definition of B that $B_p(w)=1$, in spite of the fact that w is not in $H(p)$, that is the strictly complete group B is not p-complete, showing the necessity of the condition (p).

Assume conversely that (p) is satisfied, and that B is a strictly complete group of C-characters of G. We denote by V the subgroup of all the elements in G which are mapped upon 1 by every Cp-character in B. Then $H(p) \leqq V \leqq G_{Cp}$, since the principal genus is in B. Suppose now that w is an element in G_{Cp} whose order modulo $H(p)$ is p. Then it follows from condition (p) that w is in $H(p)G_C$. It is a consequence of Theorem III.2.1(a) that $H(p)$ contains all the elements in G_C whose orders are prime to p; and hence we may represent w in the form $w=xy$ for x in $H(p)$ and y an element of order a power of p in G_C; and it follows from the C-completeness of G that y is an element of order p, since the order of xy modulo $H(p)$ is just p, and since $w \equiv y$ modulo $H(p)$. The group B is complete and contains therefore a C-character b such that $b(y) \neq 1$. Since b induces a C_0-character in G_C, and since y is an element of order p in G_C, it follows that $b(y)$ is of order p. The group B contains the C-character $b_1=b^{mp^{-m(p)}}$ which maps every element in G upon an element in $E(p)$; that is b_1 is a Cp-character and belongs to B_p. But $b_1(w)=b_1(y)$ is of order p and therefore different from 1, since the exponent $mp^{-m(p)}$ is prime to p. Thus $B_p(w) \neq 1$ and w does not belong to V. The subgroup $V/H(p)$ of $G_{Cp}/H(p)$ consequently does not contain elements of order p; and this implies $V=H(p)$, since $G_{Cp}/H(p)$ is of order a power of p. Hence B is p-complete, as was to be shown.

EXAMPLE. Every prime number p is a divisor of $q-1$ for q a suitable prime([30]); and we denote by E the cyclic group of order p^2q. There exists an automorphism g of the group E which induces both in $E(p)$ and in $E(q)$ an automorphism of order p. If G is the cyclic group of order p, generated by g, and if C is the isomorphism which maps the element g in G upon the automorphism g of E, then $G_C=G_{Cp}=G_{Cq}=1$ so that G is C-complete and so that condition (p) is satisfied. Denote by B the group of all the Cq-characters of G. Clearly B is a complete group of C-characters, since B is a complete group of Cq-characters. But $B_p=1$, showing that B is not p-complete.

This shows that the hypothesis of strict completeness occurring in Theorem IV.4.1 cannot be omitted.

REMARK. If the principal genus of G contains every C-character of G which maps G_C upon 1, then we infer $H(p)G_C=G_{Cp}$ from Theorem II.3.2 and this condition certainly implies condition (p) of Theorem IV.4.1.

COROLLARY IV.4.2. *The C-character group of G is the only strictly complete group of C-characters of G if, and only if, condition* (p) *of Theorem* IV.4.1 *is satisfied for every prime p, and $H(2)G_{C2}^2<H(2)G'$, in case C is singular.*

Proof. Suppose first that the C-character group of G is the only strictly complete group of C-characters of G. This implies the p-completeness of every strictly complete group of C-characters of G, proving the necessity of (p). It implies furthermore that the $C2$-character group of $G/H(2)$ is the only strictly complete group of $C2$-characters of G. If C is singular, then we infer from Theorem IV.2.1 that G/G_{C2} is not cyclic; and hence it follows from Theorem IV.3.8 that $H(2)G_{C2}^2<H(2)G'$.

Suppose conversely that the conditions are satisfied, and that B is a strictly complete group of C-characters of G. Then we infer from Theorem IV.4.1 that $H(p)$ is the set of all the elements in G, mapped upon 1 by the group B_p of the Cp-characters in B. If p is odd, or if $C2$ is not singular, then we infer from Theorem IV.2.1 that B_p contains every Cp-character of $G/H(p)$ and consequently of G. If $C2$ is singular, then G/G_{C2} is not cyclic, since otherwise $H(2)G_{C2}^2=H(2)G'$. Thus we may infer from our hypothesis and from Theorem IV.3.8 that B_2 contains every $C2$-character of $G/H(2)$ and of G, showing that B is the full C-character group of G.

In the following investigations we shall make use of the subgroups $\overline{H}(p)$. Each such group is defined as the cross-cut of all the subgroups $H(q)$ for $q\neq p$; and this is equivalent to saying $\overline{H}(p)$ is the set of all the elements x in G which are mapped upon elements in $E(p)$ by every C-character of G. If G is C-complete, then the cross-cut of $H(p)$ and $\overline{H}(p)$ is 1; and it follows from the fact that every C-character induces a C_0-character in the abelian group G_C that the cross-cut of G_C and $H(p)$ consists of all the elements in G_C whose

([30]) As follows from Dirichlet's theorem.

order is prime to p whereas the cross-cut of G_C and $\overline{H}(p)$ consists just of the elements of order a power of p in G_C.

THEOREM IV.4.3. *If G is C-complete, if p is a prime such that G/G_{Cp} is cyclic*(31), *then the following conditions are necessary and sufficient for p-completeness of every complete group of C-characters of G:*

(p*) *$H(p)\overline{H}(p)$ contains every element in G whose order modulo $H(p)$ is exactly p.*

(p**) *If the order of G/G_{Cp} is prime to p, and if x is an element of order a prime, not p, in $G/H(p)$, then the centralizer*(32) *of x in $G/H(p)$ contains an element different from 1 in $(H(p)\overline{H}(p))/H(p)$.*

REMARK. The condition (p) of Theorem IV.4.1 implies that elements of order p in $G_{Cp}/H(p)$ are in $(H(p)\overline{H}(p))/H(p)$, since $G_C \leqq H(p)\overline{H}(p)$; and this condition (p) is a consequence of the condition (p*) above, since $\overline{H}(p)$ is part of the cross-cut of all the G_{Cq} for $q \neq p$, and since G_C is exactly the cross-cut of all the G_{Cr}.

Proof. We precede our arguments by the proofs of some lemmas.

(1) *If the element x in $G/H(p)$ is of order a prime, not p, then the centralizer of x in $G/H(p)$ is a cyclic group T such that $G/H(p) = T(G_{Cp}/H(p))$ and such that the cross-cut of T and $G_{Cp}/H(p)$ is* 1.

We note first that $G/H(p)$ is Cp-complete. Hence $G_{Cp}/H(p)$ is, by Theorem III.1.2, a p-group which certainly does not contain the element x. It is a consequence of II.1.A that the automorphism of $E(p)$ induced by x has no fixed elements except 1; and hence it follows from Theorem III.1.2 and the Cp-completeness of $G/H(p)$ that x does not commute with any element in $G_{Cp}/H(p)$ except 1; that is, the cross-cut of T and $G_{Cp}/H(p)$ is 1.

Since G/G_{Cp} is cyclic, there exists an element g in $G/H(p)$, generating $G/H(p)$ modulo $G_{Cp}/H(p)$. The order of G/G_{Cp} is divisible by the order r of x so that the order of the automorphism g of $E(p)$ is divisible by the prime $r \neq p$; and it follows from II.1.A that the integer $g-1$ is prime to p. A certain power g^n of g induces in $E(p)$ the same automorphism as x; and we may assume without loss of generality that g has been chosen in such a fashion that n is a divisor of the order of the automorphism g of $E(p)$, that is, that nr is the order of the automorphism g of $E(p)$. Then $x' = xg^{-n}$ is an element in $G_{Cp}/H(p)$. Since g^{1-n}, $g-1$ and $g^n-1 \equiv x-1$ modulo $p^{m(p)}$ are all integers prime to p, it follows that $1+g^{-1}+\cdots+g^{1-n} \equiv g^{1-n}(1+g+\cdots+g^{n-1}) \equiv g^{1-n}(g^n-1)/(g-1)$ modulo $p^{m(p)}$ is an integer prime to p; and hence there exists an element s in $G_{Cp}/H(p)$ such that $s^{1+g^{-1}+\cdots+g^{1-n}} = x'$. But then we deduce from Theorem III.1.2 and the Cp-completeness of $G/H(p)$ that

(31) This hypothesis is automatically satisfied if p is odd. The case excluded by this hypothesis is treated in Theorem IV.4.4 below.

(32) The centralizer of the element x in the group K consists of all the elements z in K such that $zx = xz$.

$$(sg)^n = s^{1+g^{-1}+\cdots+g^{1-n}}g^n = x'g^n = xg^{-n}g^n = x.$$

Clearly $sg=t$ belongs to T; and from $g\equiv t$ modulo $G_{Cp}/H(p)$ we infer that $G/H(p)=T(G_{Cp}/H(p))$. The cyclicity of T is a consequence of the facts that the cross-cut of T and $G_{Cp}/H(p)$ is 1, and that G/G_{Cp} is cyclic. This completes the proof of (1).

(2) *If x is an element of order p in $G/H(p)$, but not in $G_{Cp}/H(p)$, then every C_0-character of $G_{Cp}/H(p)$ is induced by some Cp-character of $G/H(p)$ which maps x upon* 1.

The order of G/G_{Cp} is of the form pn; and because of the cyclicity of G/G_{Cp} there exists an element g in $G/H(p)$ such that g^n and x induce the same automorphism in $E(p)$ and such that g generates $G/H(p)$ modulo $G_{Cp}/H(p)$. Clearly $x'=xg^{-n}$ is an element in $G_{Cp}/H(p)$. Then we infer from the Cp-completeness of $G/H(p)$ and Corollary I.3.2 that

$$1 = x^p = (x'g^n)^p = x'^{1+g^{-n}+\cdots+g^{-n(p-1)}}g^{np}$$

or $g^*=g^{np}=x'^{-(1+g^{-n}+\cdots+g^{-n(p-1)})}$. We note that g^* generates the cross-cut $\{g\}_{Cp}$ of $G_{Cp}/H(p)$ and of the cyclic group $\{g\}$, generated by g. Since $G/H(p)$ is Cp-complete, so is its subgroup $\{g\}$; and it follows from Theorem I.3.1 that every C_0-character of $\{g\}_{Cp}$ is induced by some Cp-character of $\{g\}$. Furthermore we have

$$\begin{aligned} 1 + g^{-n} + \cdots + g^{-n(p-1)} &\equiv g^{-n(p-1)}(1 + g^n + \cdots + g^{n(p-1)}) \\ &\equiv g^n(1 + g^n + \cdots + g^{n(p-1)}) \text{ modulo } p^{m(p)}, \end{aligned}$$

since $g^{np}\equiv 1$ modulo $p^{m(p)}$.

Since the C-character group of a cyclic group is by Lemma III.1.1 itself a cyclic group, and since $\{g\}$ is Cp-complete, there exists a Cp-character d of $\{g\}$ with the property $d(g^i)=1$ implies $g^i=1$. Since x' is in $G_{Cp}/H(p)$, and since Cp-characters are C_0-characters in $G_{Cp}/H(p)$, we find

$$d(x'^{-g^n(1+g^n+\cdots+g^{n(p-1)})}) = d(g^*) = d(g^{np}) = d(g^n)^{1+g^n+\cdots+g^{n(p-1)}}.$$

Remembering that np is the order of the automorphism g of $E(p)$ one readily deduces from the last identity that the order of x' is a factor of the order of $d(g^n)$.

If now f is any C_0-character of $G_{Cp}/H(p)$ in $E(p)$, then there exists an integer t such that $d(g^n)^t=f(x'^{-g^n})$, since the order of $f(x'^{-g^n})$ is a divisor of the order of x'; and we verify that

$$\begin{aligned} d(g^*)^t &= d(g^{np})^t = d(g^n)^{t(1+g^n+\cdots+g^{n(p-1)})} \\ &= f(x'^{-g^n})^{1+g^n+\cdots+g^{n(p-1)}} = f(x'^{-g^n(1+g^n+\cdots+g^{n(p-1)})}) \\ &= f(g^*). \end{aligned}$$

Hence we infer from Lemma I.3.4 the existence of a Cp-character f_1 of $G/H(p)$ which induces d^t in $\{g\}$ and f in $G_{Cp}/H(p)$; and this character f_1 satisfies $f_1(x)=f_1(x'g^n)=f_1(x')^{g^n}f_1(g^n)=f(x'^{g^n})d(g^n)^t=f(x'^{g^n})f(x'^{-g^n})=1$, as was to be shown.

Suppose now that every complete group of C-characters of G is p-complete. Then we denote by B_p' the group of C-characters which is generated by all the Cq-characters for $q\neq p$.

Assume first that the element x in $G/H(p)$ is of order p. If x were in $G_{Cp}/H(p)$, then we inferred from Theorem IV.4.1(p) that x is in $G_CH(p)/H(p)\leqq(\overline{H}(p)H(p))/H(p)$. Thus we consider the case where x is not in $G_{Cp}/H(p)$. Then we denote by B_p the group of all the Cp-characters of G (or $G/H(p)$) which map x upon 1; and we put $B=B_p'B_p$. This group is certainly not p-complete, as B_p maps upon 1 all the elements in the subgroup $\{x\}$ of order p of $G/H(p)$. If y is an element in $G/H(p)$, but not in the subgroup generated by x and $G_{Cp}/H(p)$, then there exists[33] a Cp-character in the principal genus of $G/H(p)$ which maps x, but not y, upon 1; and if y belongs to the subgroup generated by x and $G_{Cp}/H(p)$, though y does not belong to $\{x\}$, then we infer from (2) the existence of a Cp-character in B_p which does not map y upon 1. This shows that $\{x\}$ is the subgroup of those elements in $G/H(p)$ which are mapped upon 1 by every Cp-character in B_p. It is a consequence of the definition of B_p' and $\overline{H}(p)$ that $\overline{H}(p)$ is the set of all the elements in G mapped upon 1 by all the C-characters in B_p'; and hence it follows that the set of elements mapped upon 1 by all the C-characters in B is just the cross-cut X^* of $\overline{H}(p)$ and the subgroup X of G which contains $H(p)$ and satisfies $\{x\}=X/H(p)$. Since B is not p-complete, it follows from our present hypothesis that $X^*\neq 1$. Since G is C-complete, it follows that the cross-cut of $H(p)$ and $\overline{H}(p)$ is 1; and since $X/H(p)$ is of order p, we deduce that $X\leqq H(p)\overline{H}(p)$, showing the necessity of condition (p*).

To show the necessity of condition (p**) we consider an element x in $G/H(p)$ whose order is a prime $r\neq p$. Then we infer from (1) the existence of an element g in $G/H(p)$ such that g generates $G/H(p)$ modulo $G_{Cp}/H(p)$ and such that $g^n=x$ where nr is the order of the automorphism g of $E(p)$. We consider the group B_p of all the Cp-characters of G or $G/H(p)$ which map g upon 1. It is a consequence of (1) and of Lemma I.3.4 that every C_0-character of $G_{Cp}/H(p)$ in $E(p)$ is induced by one and only one Cp-character in B_p; and this shows that the group $\{g\}$ generated by g is the subgroup of all the elements in $G/H(p)$ which are mapped upon 1 by every Cp-character in B_p. Denote by B the product[34] of B_p' and B_p. Then the group of elements mapped upon 1 by all the C-characters in B is the cross-cut X^* of $\overline{H}(p)$ and

(33) Take an element v of order p in $E(p)$. Then $v^{1-x}=1$, though $v^{1-y}\neq 1$, since the order of y modulo G_{Cp} is neither 1 nor p, as G/G_{Cp} is cyclic.

(34) Where B_p' denotes, as before, the group generated by all the Cq-characters for q any prime not p, that is B_p' consists of all the C-characters of order prime to p.

of the subgroup X of G which contains $H(p)$ and satisfies $X/H(p)=\{g\}$. Since B is not p-complete, it is not complete, showing that $X^*\neq 1$; and noting that $\{g\}$ is exactly the centralizer of x in $G/H(p)$ (cp. (1)!), we deduce the necessity of conditon (p**).

Assume now that conditions (p*) and (p**) are both satisfied by G, C, and p, and that $B=B_p'B_p$ is a complete group of C-characters of G where B_p consists of Cp-characters of G only, whereas B_p' is generated by Cq-characters of G with $q\neq p$. This implies in particular that $B_p'(\overline{H}(p))=1$; and we deduce from the completeness of B that there exists to every element, not 1, in $\overline{H}(p)$ a Cp-character in B_p which does not map it upon 1.

Consider an element x in G whose order modulo $H(p)$ is p. Then we infer from condition (p*) that x is in $H(p)\overline{H}(p)$; and one deduces from the remark made at the end of the last paragraph the existence of a Cp-character in B_p which does not map x upon 1.

Next we consider([35]) an element x whose order modulo $H(p)$ is a prime $r\neq p$. Then we infer from (1) that the centralizer T of $H(p)x$ in $G/H(p)$ is a cyclic group, generated by an element g, such that the cross-cut of T and $G_{Cp}/H(p)$ is 1, and such that $G/H(p)=T(G_{Cp}/H(p))$. If p is a divisor of the order of G/G_{Cp}, then p is a divisor of the order of T; and it follows from (p*) that the cross-cut of T and $(\overline{H}(p)H(p))/H(p)$ is different from 1. If p is not a divisor of the order of G/G_{Cp}, then we infer from condition (p**) that the cross-cut of T and $(\overline{H}(p)H(p))/H(p)$ is different from 1. Denote by T^* the cross-cut of T and $(\overline{H}(p)H(p))/H(p)$. Then we have seen that T^* is a cyclic group different from 1; and there exists an element t in $\overline{H}(p)$ such that $H(p)t$ generates T^*. We note that $t\neq 1$. Since t is in $\overline{H}(p)$, it follows that $B_p'(t)=1$; and we infer from the completeness of B the existence of a Cp-character b in B_p such that $b(t)\neq 1$. It is a consequence of Theorem II.2.1 that every Cp-character of T belongs to the principal genus; and hence there exists an element v in $E(p)$ such that $b(y)=v^{1-y}$ for y in T. From $1\neq b(t)=v^{1-t}$ we infer $v\neq 1$; and since $x-1$ is by II.1.A prime to p, it follows that $b(x)=v^{1-x}\neq 1$. Thus we have shown that no element of order a prime in $G/H(p)$ is mapped upon 1 by every Cp-character in B (or B_p); and from this fact one deduces the p-completeness of B.

In Theorem IV.4.3 we discussed only those primes p for which G/G_{Cp} is cyclic; and this is always the case if p is odd. Thus we have to discuss the situation in case G/G_{C2} is not cyclic. Then there exists a coset Z_0 of G/G_{C2} all of whose elements induce in $E(2)$ the inversion. Apart from Z_0 there are two further cosets of order 2 in G/G_{C2}; and it follows from II.1.B that the integers corresponding under $C2$ to these two cosets are of the form $\pm 1+2^{m(2)-1}$ modulo $2^{m(2)}$. We denote by K_0 the uniquely determined coset of order 2 of G/G_{C2} to whose elements there corresponds under $C2$ the in-

([35]) Since $G/H(2)$ is a 2-group, the existence of such an element x implies that p is odd.

teger $1+2^{m(2)-1}$ modulo $2^{m(2)}$; and we denote by $K=K(G, C)$ the subgroup of G whose elements are either in G_{C2} or in K_0. In case the order of G/G_{C2} is divisible by 8 it is possible to characterize K as the subgroup between G_{C2} and G^2G_{C2} with the property that K/G_{C2} is of order 2.

THEOREM IV.4.4. *If G is C-complete, and if G/G_{C2} is not cyclic, then the following conditions are necessary and sufficient for 2-completeness of every complete group of C-characters of G:*

(i) *$H(2)\overline{H}(2)$ contains every element in the subgroup K whose order modulo $H(2)$ is 2([36]).*

(ii) *If G/G_{C2} is of order 4, then $\overline{H}(2)H(2)$ contains the following elements:*

(ii′) *Every element x in G whose order modulo $H(2)$ is 2 and which does not induce the inversion in $E(2)$, and*

(ii″) *Every element z in G which induces the inversion in $E(2)$ and which satisfies $(zy)^2\equiv y^{2^{m(2)-1}}$ modulo $H(2)$ for every y in K, y not in G_{C2}.*

Proof. We start by proving some lemmas.

(1) *If S is a subgroup of $K/H(2)$, then every $C2$-character of S is induced by some $C2$-character of G.*

It is obvious that the group of all the $C2$-characters of G induces in S a complete group of $C2$-characters. But the only complete group of $C2$-characters of $S\leqq K/H(2)$ is the group of all the $C2$-characters of S, as follows from Theorem IV.2.2, if S is not part of $G_{C2}/H(2)$, and from Theorem I.1.4, if $S\leqq G_{C2}/H(2)$.

(2) *If S is a subgroup of $K/H(2)$, and if $P(S)$ is the group of $C2$-characters of G, mapping S upon 1, then there exists to every element in $K/H(2)$, not in S, a $C2$-character in $P(S)$ which does not map it upon 1.*

If P_2 is the group of all the $C2$-characters of G, then we infer from (1) that $P_2/P(S)$ is essentially the same as the group of all the $C2$-characters of S. It follows from Theorem II.5.2 that the orders of S and of its group $P_2/P(S)$ of $C2$-characters are equal. If S'' is the group of elements in $K/H(2)$ mapped upon 1 by all the $C2$-characters in $P(S)$, then $S\leqq S''$ and $P(S)=P(S'')$ showing that S and S'' are of equal order; and this implies the desired equality of S and S''.

Assume now that every complete group of C-characters of G is 2-complete. Consider an element x of order 2 in $K/H(2)$; and denote by B_2 the group of all the $C2$-characters of G which map x upon 1, by Q the group of all the C-characters of odd order of G, and by B the product of Q and B_2. The group of elements in G, mapped upon 1 by Q, is $\overline{H}(2)$; and the group of elements in G mapped upon 1 by B_2 is, by (2), the group X, generated by the elements in the coset x of $K/H(2)$. Since B is obviously not 2-complete, it follows that B

([36]) Thus $H(2)\ \overline{H}(2)$ contains every element in G_{C2} whose order modulo $H(2)$ is 2; and it may happen that these are the only elements in K whose orders modulo $H(2)$ are 2.

is not complete; and this implies that the cross-cut of X and $\overline{H}(2)$ is different from 1. Since $X/H(2)$ is of order 2, and since the cross-cut of $H(2)$ and $\overline{H}(2)$ is 1 as a consequence of the C-completeness of G, it follows now that $X \leqq H(2)\overline{H}(2)$, proving the necessity of condition (i).

Suppose now that the order of G/G_{C2} is 4. If x is an element of order 2 in $G/H(2)$ which does not induce the inversion in $E(2)$, then it is a consequence of condition (i), already proven, that x is in $(\overline{H}(2)H(2))/H(2)$, provided x is in $K/H(2)$. If x is not in $K/H(2)$, then the integer corresponding to x under $C2$ is of the form $x \equiv -1+2^{m(2)-1}$ modulo $2^{m(2)}$. We denote by B_2 the group of all the $C2$-characters of G which map x upon 1; and by W the group, generated by adjoining x to $G_{C2}/H(2)$. From $x^2=1$ and Lemma I.3.4 we infer that every C_0-character of $G_{C2}/H(2)$ is induced by a $C2$-character of W which maps x upon 1; and it follows from Corollary III.1.3 and the fact that W is of index 2 in $G/H(2)$ that every $C2$-character of W is induced by a $C2$-character of $G/H(2)$. Since B_2 contains the $C2$-character which maps W upon 1 and the elements in $G/H(2)$ which are not in W upon the element of order 2 in $E(2)$, it follows that only elements in W are mapped upon 1 by B_2. Since every C_0-character of $G_{C2}/H(2)$ is induced by a $C2$-character in B_2, and since $G/H(2)$ is $C2$-complete, it follows that elements not equal to 1 in $G_{C2}/H(2)$ are not mapped upon 1 by every $C2$-character in B_2. Thus it follows that x and 1 are the only elements mapped upon 1 by every $C2$-character in B_2. Denote now by Q the group of all the C-characters of odd order of G, and by B the product of Q and B_2. If we denote by X the subgroup consisting of the elements in $H(2)$ and in the coset x, then the cross-cut of X and $\overline{H}(2)$ is the set of elements mapped upon 1 by B. Since B is not 2-complete, it follows that B is not complete and that therefore the cross-cut of X and $\overline{H}(2)$ is not 1. Since $X/H(2)$ is of order 2, this implies $X \leqq H(2)\overline{H}(2)$, that is, we have verified the necessity of condition (ii').

Suppose again that G/G_{C2} is of order 4. Then G may be generated by adjoining to K the elements in G which induce the inversion in $E(2)$. Assume furthermore that the element z in $G/H(2)$ induces the inversion in $E(2)$ and satisfies $(zy)^2=y^{2^{m(2)-1}}$ for every y in $K/H(2)$ which is not in $G_{C2}/H(2)$. Then it follows from Lemma IV.3.1(c) that every $C2$-character f of $G/H(2)$ satisfies

$$\begin{aligned} f(z)^{1-y} &= f([z, y])^y f(y)^2 = f(zyzyy^{-2})^y f(y)^2 = f(y^{2^{m(2)-1}-2})^y f(y)^2 \\ &= f(y^2)^{y(2^{m(2)-2}-1)} f(y)^2 = f(y)^{(1+y)y(2^{m(2)-2}-1)} f(y)^2 \\ &= f(y)^{(y+1)(2^{m(2)-2}-1)} f(y)^2 \\ &= f(y)^{2(1+2^{m(2)-2})(2^{m(2)-2}-1)} f(y)^2 = f(y)^{-2} f(y)^2 = 1; \end{aligned}$$

and it follows from Corollary III.1.3 that every $C2$-character of $K/H(2)$ is induced by a $C2$-character of $G/H(2)$ which maps z upon 1. If we denote by B_2 the group of all the $C2$-characters of G which map the coset z of $G/H(2)$

upon 1, then it is readily seen that the subgroup X, generated by the elements in the coset z and containing $H(2)$, consists exactly of the elements in G mapped upon 1 by the $C2$-characters in B_2, since $K/H(2)$ is $C2$-complete. Denote by Q again the group of C-characters of odd order of G. Then $\overline{H}(2)$ is the set of elements mapped upon 1 by Q. The group $B=QB_2$ is not 2-complete; and thus it follows that B is not complete. This implies that the cross-cut of $\overline{H}(2)$ and X is different from 1; and we deduce $X \leqq H(2)\overline{H}(2)$ from the fact that $X/H(2)$ is of order 2, proving the necessity of condition (ii'').

We assume now conversely that the conditions (i) and (ii) are satisfied by G, C, and that B is a complete group of C-characters of G. This group B is the direct product of the groups B_2 and P where P consists of the C-characters of odd order in B whereas B_2 is the subgroup of all the $C2$-characters in B. We denote by R a group between K and G whose elements do not induce the inversion in $E(2)$ and whose index in G is 2. If x is an element of order 2 in $R/H(2)$, then x is in $K/H(2)$, since $R/H(2)$ is cyclic. Consequently we may deduce from condition (i) that x is in $(\overline{H}(2)H(2))/H(2)$. There exists therefore an element $w \neq 1$ in $\overline{H}(2)$ such that $x=H(2)w$. From the completeness of B we infer the existence of a $C2$-character b in B_2 which does not map w upon 1. Hence $b(x)=b(w)\neq 1$, showing that B_2 induces a complete group of $C2$-characters in $R/H(2)$. Thus we may deduce from Theorem IV.2.2 that every $C2$-character of $R/H(2)$ is induced by some $C2$-character in B_2.

If s is an element of order 2 in $G/H(2)$, and if s is in $R/H(2)$, then we have already shown that $B_2(s)\neq 1$. If s is not in $R/H(2)$, then we distinguish two cases.

Case 1. *s does not induce an inversion in $E(2)$.* Then there exists an element g generating $R/H(2)$ modulo $G_{C2}/H(2)$ and an inversion z in $G/H(2)$ such that $s=zg^{2^i}$ for some i. Since s is not an inversion, g^{2^i} is not in $G_{C2}/H(2)$. Since s is of order 2, we find that

$$1 = s^2 = (zg^{2^i})^2 = zg^{2^i}zg^{-2^i}g^{2^{i+1}} = [z, g^{2^i}]g^{2^{i+1}},$$

$$g^{2^{i+1}} = [g^{2^i}, z],$$

since inversions are induced in $E(2)$ by elements of order 2 in $G/H(2)$, as follows from Theorem III.1.2. Thus it follows that the automorphism g^{2^i} of $E(2)$ is of order 2, that g^{2^i} is therefore in $K/H(2)$, and that the integer corresponding under $C2$ to the element g^{2^i} in $K/H(2)$ is $g^{2^i}\equiv 1+2^{m(2)-1}$ modulo $2^{m(2)}$.

Every $C2$-character of $R/H(2)$ is induced by some $C2$-character in B_2. Thus there exists a $C2$-character f in B_2 such that $f(t)=e^{1-t}$ for t in $R/H(2)$ where e is an element generating $E(2)$. If $0<i$, then the order of $g^{2^{i+1}}$ does not exceed $2^{m(2)-i-1}$ which is a divisor of $2^{m(2)-2}$; and thus it follows from Corollary III.1.3 that $f(z)$ is of the form $e^{2+fe^{2^{m(2)-2}}}$, generates $E(2)^2$. Consequently we find that $f(zg^{2^i})=f(z)e^{2^i}f(g^{2^i})=f(z)^{1+2^{m(2)-1}}e^{2^{m(2)-1}}\neq 1$, since the second factor

is of order 2, the first is of order $2^{m(2)-1}$, that is $B_2(s) \neq 1$. If, however, $i=0$, then g^2 is in $G_{C2}/H(2)$ and the order of G/G_{C2} is 4. In this case we infer from condition (ii') that s is in $(\overline{H}(2)H(2))/H(2)$; and there exists an element s'' in $\overline{H}(2)$ such that $s=H(2)s''$. Since B is complete, and since $P(\overline{H}(2))=1$, there exists a $C2$-character d in B_2 such that $d(s)=d(s'')\neq 1$; and we have shown again that $B_2(s)\neq 1$.

Case 2. An inversion is induced by s in $E(2)$. Then we consider again an element g in $R/H(2)$ which generates R modulo G_{C2}. There exists a $C2$-character f in B_2 such that $f(t)=e^{1-t}$ for t in $R/H(2)$ where the element e generates $E(2)$. It follows from Lemma IV.3.1(c) that $f(s)^{1-g}=e^{2(g-1)}$; and this implies $f(s)\neq 1$, provided the automorphism g of $E(2)$ is not of order 2; $B_2(s)\neq 1$ provided G/G_{C2} is not of order 4.

If G/G_{C2} is of order 4, then it may happen that there exists a $C2$-character f of $R/H(2)$ such that $f([s,g])^g f(g)^2 \neq 1$. This $C2$-character is induced by some $C2$-character f in B_2; and it follows from Lemma IV.3.1(c) that $f(s)^{1-g}\neq 1$ so that again $B_2(s)\neq 1$.

Suppose finally that $1=f([s,g])^g f(g)^2$ for every $C2$-character f of $R/H(2)$. Then we note that the integer corresponding to g under $C2$ is $g\equiv 1+2^{m(2)-1}$ modulo $2^{m(2)}$, since g induces an automorphism of order 2 in $E(2)$, and since g is in $R=K$. Consequently we find:

$$\begin{aligned} f((sg)^2 g^{-2^{m(2)-1}}) &= f(sgsg^{-1}g^{2-2^{m(2)-1}}) = f([s,g])f(g^{2-2^{m(2)-1}}) \\ &= f([s,g])f(g^2)^{1-2^{m(2)-2}} \qquad \text{(since } g^2 \text{ is in } G_{C2}/H(2)\text{)} \\ &= f([s,g])f(g)^{(1+g)(1-2^{m(2)-2})} \\ &= f([s,g])f(g)^2 = f([s,g])^g f(g)^2 = 1, \end{aligned}$$

since $f([s,g])$ is of an order dividing $2^{m(2)-1}$, and we infer from the $C2$-completeness of $G/H(2)$ that $R/H(2)$ is $C2$-complete, and that therefore $(sg)^2=g^{2^{m(2)-1}}$. Since g may be any element in $R/H(2)=K/H(2)$ which is not in $G_{C2}/H(2)$, we may apply condition (ii); and it follows that s is in $(\overline{H}(2)H(2))/H(2)$; and there exists an element s_0 in $\overline{H}(2)$ such that $s=H(2)s_0$. Since B is complete, we have $B(s_0)\neq 1$. Since s_0 is in $\overline{H}(2)$, we have $Q(s_0)=1$. Hence there exists a $C2$-character b in B_2 such that $b(s)=b(s_0)\neq 1$, proving again that $B_2(s)\neq 1$.

Thus we have finally shown that $B_2(s)\neq 1$ for every s in G whose order modulo $H(2)$ is exactly 2; and that proves the 2-completeness of B, as was to be shown.

For the enunciation of the next theorem we need a restatement of a concept introduced before (cp. Lemma IV.3.4!): If G/G_{C2} is not cyclic, then there exists one and only one subgroup $A=A(G,C)$ which contains G_{C2}, is of index 2 in G, does not contain any element inducing the inversion in $E(2)$, and contains an element such that the integer corresponding to it under C_2 is congruent to -1 modulo 4.

THEOREM IV.4.5. *If G is C-complete, then the following conditions are necessary and sufficient for the full C-character group of G to be the only complete group of C-characters of G:*

(a) *If G/G_{Cp} is cyclic, then $\overline{H}(p)H(p)$ contains every element in G whose order* modulo $H(p)$ *is p.*

(b) *If G/G_{Cp} is of order prime to p, and if the element x in G is* modulo $H(p)$ *of order a prime number different from p, then $\overline{H}(p)H(p)$ contains an element y which is not in $H(p)$, but satisfies $xy \equiv yx$* modulo $H(p)$.

(c) *If G/G_{C2} is not cyclic, then $\overline{H}(2)H(2)$ contains every element in the subgroup K (introduced in Theorem* IV.4.4) *whose order* modulo $H(2)$ *is* 2.

(d) *If p is odd and $G \neq G_{Cp}$, then either $m(p)=1$ or the order of G/G_{Cp} is divisible by p.*

(e) *If G/G_{C2} is of order* 2, *and if there exists an element in G to which there corresponds under $C2$ an integer congruent to* -1 modulo 4, *then $H(2)G_{C2}^2 < H(2)G^2$.*

(f) *If G/G_{C2} is not cyclic, and if the order of G/G_{C2} is divisible by* 8, *then $H(2)\,G'$ is not part of $A^2H(2)$.*

(g) *If G/G_{C2} is not cyclic, but of order* 4, *then neither $H(2)G'$ nor $H(2)A^2$ is equal to $H(2)G_C^2$.*

Proof. It is readily seen that the following two conditions are necessary and sufficient for the full C-character group of G to be the only complete group of C-characters of G:

(A) Every complete group of C-characters of G is p-complete, for every prime p.

(B) The full Cp-character group of G (or of $G/H(p)$) is the only complete group of Cp-characters of $G/H(p)$, for every prime p.

It is an immediate consequence of Theorems IV.2.2 and IV.3.7 that the conditions (d) to (g) of the theorem are equivalent to the property (B); and it is a consequence of Theorems IV.4.3 and IV.4.4 that (A) implies conditions (a) to (c). If conversely all the conditions (a) to (g) are satisfied by G and C, then the conditions (p*) and (p**) of Theorem IV.4.3 are consequences of (a) and (b) respectively; and condition (i) of Theorem IV.4.4 is an immediate consequence of the present condition (c). Condition (ii'') of Theorem IV.4.4 is satisfied, since we may infer from (g) the impossibility of the existence of an element z meeting the requirements of condition (ii'') of Theorem IV.4.4 considering that $2 < m(2)$, if G/G_{C2} is not cyclic. Condition (ii') of Theorem IV.4.4 is a consequence of condition (c) of the present theorem, since elements of order 2 in $G/H(2)$ which do not induce the inversion in $E(2)$ belong by Theorem IV.4.5(i) and Lemma IV.3.4 to $K/H(2)$, if the order of G/G_{C2} is 4 and G/G_{C2} is not cyclic. Thus it follows from Theorems IV.4.3 and IV.4.4 that condition (A) is a consequence of the conditions (a) to (g); and this completes the proof.

CHAPTER V. DUALITY

A *projectivity* of the group S upon the group T is a biunivoque correspondence which maps the set of all the subgroups of S upon the set of all the subgroups of T and which has the property of preserving the relation "$\leqq$." A *duality* of the group S upon the group T is a biunivoque correspondence which maps the set of all the subgroups of S upon the set of all the subgroups of T and which has the property of inverting the relation "$\leqq$." If the product of two dualities exists, then it is a projectivity.

If G is a group and C a homomorphism of G into the group of automorphisms of the cyclic group E of order m, then we denote the *C-character group of* G by $L=L(G, C)$. This group L is an abelian group the orders of whose elements are divisors of m; and thus it follows from Theorems I.1.3 and I.1.4 that L and the group L_0 of its C_0-characters in E are isomorphic groups; and that there exists a duality between L and L_0. Thus the problems of constructing a duality between G and L and that of constructing a projectivity between G and L_0 are equivalent problems.

V.1. **The natural projectivity between G and L_0.** Throughout this section we make use of the notations (II.3.*). In the proof of Theorem IV.1.2 we have introduced the operator F_g, defined by the equation $F_g(f)=f(g)$ for g an element in G and f a C-character of G.

LEMMA V.1.1. *Suppose that G is C-complete.*

(a) *F_g is a C_0-character of the C-character group of G, that is F_g is, for every g in G, an element in L_0.*

(b) *If s and t are elements in G such that $F_s=F_t$, then $s=t$.*

(c) *If S is a subset of G such that the set F_S is a subgroup of L_0, then S is a subgroup of G.*

REMARK. Note that F_G need not be a subgroup of L_0, though F_G is always a subset of L_0. In this respect see Theorem V.1.2.

Proof. Statements (a) and (b) are easily verified; see for example (I.3.5.1) and (I.3.5.2). Suppose now that S is a subset of G such that F_S is a subgroup of L_0, and that s, t are elements in S. Then the characters F_s, F_t, F_s^{-1} and $F_s^{-1}F_t$ belong to F_S. If f is any C-character of G, then $F_s^{-1}F_t(f)=(F_s(f))^{-1}F_t(f)=f(s)^{-1}f(t)=f(st)=F_{st}(f)$; and it follows from (b) that st is in S, proving (c).

THEOREM V.1.2. *If G is C-complete, then each of the following properties implies the others:*

(1) *The set F_G is a group of C_0-characters of $L(G, C)$.*

(2) *C is regular.*

(3) *F_S is a group of C_0-characters of $L(G, C)$ if, and only if, S is a subgroup of G.*

REMARK. It is readily deduced from Lemma V.1.1 that F_S is a projectivity of G upon L_0 if, and only if, (3) holds true.

Proof. It is obvious that (1) is a consequence of (3). If (1) is satisfied by G, then F_G is a complete group of C_0-characters of $L(G, C)$, since $F_0(f)=1$ for every g implies $f(g)=1$ for every g, that is, implies $f=1$. Hence it follows from Theorem I.1.4 that F_G is the group L_0 of all the C_0-characters of $L(G, C)$ in E. It follows from Theorem I.1.3 and Lemma V.1.1(b) that G, $F_G=L_0$, and L are of equal order, and it follows from Corollary II.5.3(b) that (2) is a consequence of (1). Suppose finally that (2) is satisfied by G and C. Then we deduce from Lemma V.1.1(c) that we need prove only the following fact: F_S is a group, whenever S is a group. If S is a subgroup of G, then S is C-complete and satisfies the same condition (2) as G. Hence it follows from Corollary II.5.3(b) that S and its C-character group $L(S, C)$ are of equal order. Since $L(G, C)$ induces in S a complete group of C-characters, and since it follows from Theorem IV.1.2 that the C-character group of S is the only complete group of C-characters of S, we find that every C-character of S is induced by some C-character of G. If we denote by L_S the group of all the C-characters of G which map S upon 1, then it follows that $L(S, C)$ is essentially the same as $L(G, C)/L_S$; and since $L(S, C)$ and S, $L(G, C)$ and G are of equal order, we deduce that the order of L_S is the index $[G:S]$ of S in G. Denote by L_S^* the group of all the C_0-characters of $L(G, C)$ in E which map L_S upon 1. It is obvious that $F_S \leqq L_S^*$, since $f(S)=1$ and $F_S(f)=1$ are equivalent statements. It is a consequence of Theorems I.1.3 and I.1.4 that the order of L_S^* is the index of the subgroup L_S of $L(G, C)$ in $L(G, C)$. Since G and $L(G, C)$ are of equal order, and since the order of L_S is $[G:S]$, it follows that S and L_S are of equal order, proving that $F_S=L_S^*$, since F_S contains by Lemma V.1.1(b) as many elements as S. Since L_S^* is a group of C_0-characters of $L(G, C)$, it follows now that F_S is a group of C_0-characters of $L(G, C)$, as was to be shown.

THEOREM V. 1. 3. *The C-complete group G and its C-character group L are isomorphic if, and only if, G is abelian and C is regular.*

Proof. If G and L are isomorphic, then G is commutative, since L is commutative. Both groups furthermore have equal order and the regularity of C is a consequence of Corollary II.5.3.

Suppose conversely that G is commutative and that C is regular. Then it follows from Theorem V.1.2 that a projectivity of G upon L_0 is effected by mapping the subgroup S of G upon the subset F_S of L_0. Since S and F_S are groups of equal order, it may be deduced from a well known theorem([37]) that

([37]) See Rottländer, Math. Zeit. vol. 28 (1928) pp. 641–653 or R. Baer, Amer. J. Math. vol. 61 (1938) p. 30; one may verify this fact readily by showing the equality of the invariants of G and L_0.

G and L_0 are isomorphic. We infer from Theorem I.1.3 that L and L_0 are isomorphic, proving that G and L are isomorphic too.

V.2. **The natural duality between G and L.** The following procedure is customary for setting up a duality between groups and their character groups: using the notations (II.3.*) and assuming that G be C-complete, we denote by (L, S), for S a subgroup of G, the set of all the C-characters of G which map S upon 1; and we denote by (G, T), for T a subgroup of the C-character group L of G, the set of all the elements x in G, satisfying $T(x)=1$. Clearly (L, S) is a subgroup of L and (G, T) a subgroup of G; and it is readily verified that these correspondences effect dualities if, and only if,

(V.2.1) $(L, (G, T))=T$ *and* $(G, (L, S))=S$ *for every subgroup* S *and* T *of* G *and* L *respectively.*

It is our object to determine those G, C which meet the requirement (V.2.1). It is easy to see for example that the groups G, satisfying condition (3) of Theorem V.1.2, have the property (V.2.1).

THEOREM V.2.2. *If G is C-complete, then the following conditions are necessary and sufficient for G, C to satisfy* (V.2.1):

(i) *The orders of $H(p)$ and of $G/H(p)$ are relatively prime.*

(ii) *If the order of G/G_{Cp} is not a power of p, then $m(p)=1$ and the order of G/G_{Cp} is a prime (dividing $p-1$).*

(iii) *$C2$ is regular.*

Proof. Suppose first that (V.2.1) is satisfied by G and C. Denote by L_p the group of Cp-characters of G. Then $H(p)=(G, L_p)$ is nothing but a restatement of the definition of $H(p)$. We deduce from (V.2.1) that $L_p=(L, H(p))$ and that therefore the full Cp-character group L_p of $G/H(p)$ is the only complete group of Cp-characters of $G/H(p)$. Hence we infer from Theorem IV.2.2(1) that either the order of G/G_{Cp} is divisible by p or $G=G_{Cp}$ or $m(p)=1$. Denote by P_p the Cp-principal genus of G (or $G/H(p)$). Then $G_{Cp}=(G, P_p)$ and we deduce from (V.2.1) that $P_p=(L, G_{Cp})$. Applying this last result on $p=2$, we infer from Theorem II.3.2 that $C2$ cannot be singular. Every G/G_{Cp} is therefore cyclic. Since P_p is by Theorem II.2.1 a cyclic group of order a power of p, we may infer from (V.2.1) that G/G_{Cp} is a cyclic group of prime power order, containing as many subgroups as does P_p; and the necessity of the conditions (ii) and (iii) is now an immediate consequence of Theorem II.2.1.

Since L is the direct product of all the L_p, we find that $\overline{H}(p)=(G, \prod_{p\neq q}L_q)$; since the cross-cut of L_p and $\prod_{p\neq q}L_q$ is 1, then G is the product of $H(p)$ and $\overline{H}(p)$ (for every p); and this proves that G is the direct product of all the $\overline{H}(p)$. Since the orders of L_p and $\prod_{p\neq q}L_q$ are relatively prime, it follows that every subgroup T of L is the cross-cut of TL_p and of $T\prod_{p\neq q}L_q$; and hence it follows from (V.2.1) that every subgroup S of G is the product of its cross-cut with $H(p)$ and of its cross-cut with $\overline{H}(p)$, proving that the orders of

$H(p)$ and $\overline{H}(p)$ are relatively prime; and thus we have verified the necessity of (i).

Suppose conversely that the conditions (i) to (iii) are satisfied by G and C. Then we infer from (i) and Corollary II.4.4 that G is the direct product of its subgroups $\overline{H}(p)$; and that G is, for every prime p, the direct product of $H(p)$ and $\overline{H}(p)$.

Consider a subgroup S of $\overline{H}(p)$. Since every C-character of G maps $\overline{H}(p)$ into $E(p)$, it follows that $\overline{H}(p)$ and S are mapped upon 1 by every Cq-character for $q \neq p$; and since G is the direct product of $H(p)$ and $\overline{H}(p)$, it is readily seen that the C-characters of G induce in $\overline{H}(p)$ a group of Cp-characters which is essentially the same as the group L_p of the Cp-characters of G. Thus $(L, S) = (L, S)_p L_p'$ where $L_p' = \prod_{p \neq q} L_q$ and where $(L, S)_p$ is the group of Cp-characters of G which map S upon 1; and this shows that $(G, (L, S))$ is exactly the cross-cut of $(G, (L, S)_p)$ and $\overline{H}(p)$, that is $(G, (L, S))$ consists of those elements in $\overline{H}(p)$ which are mapped upon 1 by the Cp-characters in (L, S). If the order of G/G_{Cp} is a power of p, then we infer from (iii) the regularity of Cp and it follows from Theorem V.1.2 that $S = (G, (L, S))$. If the order of G/G_{Cp} is not a power of p, then it follows from (ii) that $m(p) = 1$ and that G/G_{Cp} is of order a prime $p^* \neq p$. Consequently either $S \leqq \overline{H}(p)_{Cp}$ or $\overline{H}(p) = S\overline{H}(p)_{Cp}$; in the first of these cases our contention $S = (G, (L, S))$ is an immediate consequence of Theorems I.1.4 and I.3.1; in the second of these cases we infer from Lemma I.3.4 that every C_0-character of $\overline{H}(p)_{Cp}$ in $E(p)$ which maps S_{Cp} upon 1 is induced by a Cp-character in $(L, S)_p$; and $S = (G, (L, S))$ is a consequence of Theorem I.1.4

If S is a subgroup of G, and if S_p is the cross-cut of S and $\overline{H}(p)$ then S is the direct product of the S_p, since the orders of $H(p)$ and $\overline{H}(p)$ are relatively prime by (i). If f is a C-character of G, then f is the product of uniquely determined Cp-characters f_p. Since $f_p(H(p)) = 1$, it follows that $f(S) = 1$ if, and only if, $f_p(S_p) = 1$ for every p. This shows that (L, S) is the direct product of the groups $(L, S_p)_p$ consisting of all the Cp-characters of G which map S_p upon 1. But we have shown in the preceding paragraph of the proof that $S_p = (G, L_p'(L, S_p)_p)$ and this makes the desired identity, $S = (G, (L, S))$ evident.

If T is a subgroup of the group L_p of the Cp-characters of G, then (G, T) contains $H(p)$, since $L_p(H(p)) = 1$. If f is a Cq-character of G for $q \neq p$ which maps $H(p)$ upon 1, then f maps $\overline{H}(p)$ upon 1 too; and this implies $f = 1$, since G is the direct product of $H(p)$ and $\overline{H}(p)$. Thus it follows that $(L, (G, T))$ is a group of Cp-characters of G. If the order of G/G_{Cp} is a power of p, then it follows from (iii) and Theorem V.1.2 that $T = (L, (G, T))$. If the order of G/G_{Cp} is not a power of p, then it follows from (ii) that $m(p) = 1$ and that G/G_{Cp} is of order a prime $p^* \neq p$. If the principal genus P_p of $G/H(p)$ in $E(p)$ is part of T, then (G, T) is between $H(p)$ and G_{Cp}; and it is readily verified that $T = (L, (G, T))$ by Theorems I.1.4 and I.3.1 and the fact that $(L, (G, T))$

is a group of Cp-characters. Thus we assume finally that P_p is not part of T. We note that P_p is of order p, and that every Cp-character of G different from 1 is of order p, since $E(p)$ is of order p. We wish to show that (G, T) contains an element not in G_{Cp}. Denote by g an element in $G/H(p)$, not in $G_{Cp}/H(p)$. Then g generates G modulo G_{Cp}, since G/G_{Cp} is of order a prime p^*; and it is a consequence of the Cp-completeness of $G/H(p)$ and of II.1.A and Theorem I.3.1 that p^* is the order of g. Our object is attained if g is in (G, T); and thus we assume that g is not in (G, T). We denote by T^* the subgroup of those Cp-characters in T which map g upon 1; since $E(p)$ is of order p, and since the Cp-characters of $G/H(p)$ therefore map g upon p different values, it follows that T/T^* is of order p. Since P_p is of order p, and since P_p is not part of T, it follows that the cross-cut of T and P_p is 1; and hence it follows from Theorem II.2.1(a) that different Cp-characters in T induce different C_0-characters of $G_{Cp}/H(p)$ in $E(p)$. Hence there exists an element in $G_{Cp}/H(p)$ which is mapped upon 1 by T^*, but not by T. Since the elements in $G_{Cp}/H(p)$ are of order p, this implies the existence of an element y in $G_{Cp}/H(p)$ such that the Cp-characters in T have the same values on g and on y. Consequently gy^{-1} is mapped upon 1 by the Cp-characters in T so that the elements in the coset gy^{-1} are in (G, T), though not in G_{Cp}. Denote now by w any element in (G, T), not in G_{Cp}. If f is a C-character in $(L, (G, T))$, then f is a Cp-character of G; and there follows from Theorem I.1.4 the existence of a Cp-character f^* in T which coincides with f on G_{Cp}. But both f and f^* map w upon 1; and G is generated by adjoining w to G_{Cp}. Hence $f=f^*$ is in T, proving the desired equation $T=(L, (G, T))$.

Consider now any subgroup T of L; and denote by T_p the group of the Cp-characters in T. Then T is the direct product of the subgroups T_p; and (G, T) is the cross-cut of the groups (G, T_p). Since (G, T_p) contains, as has been remarked before, the subgroup $H(p)$, it follows that (G, T_p) is the direct product of $H(p)$ and of a subgroup U_p of $\overline{H}(p)$, as G is the direct product of $H(p)$ and $\overline{H}(p)$. Since U_p is part of every $H(q)$ for $q \neq p$, it follows that U_p is part of (G, T). If f is a Cp-character in $(L, (G, T))$, than f maps both $H(p)$ and U_p upon 1 so that f belongs to $(L, (G, T_p))$. But we have shown in the preceding paragraph of the proof that $T_p=(L, (G, T_p))$ showing that f is in T; and now it is obvious that $T=(L, (G, T))$, as was to be shown.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

ON SPACES IN WHICH TWO POINTS DETERMINE A GEODESIC

BY

HERBERT BUSEMANN

The present note deals with spaces in which geodesics exist and are uniquely determined by any two of their points.

The existence of geodesics without additional topological or differentiability conditions finds its expression in the Axioms A, B, C, D below. It may be helpful to point out that a symmetric variational problem in parametric form will satisfy these axioms when the extremals are considered as geodesics. It will appear that there are only two different types of spaces in which the geodesic through any two distinct points is unique, similar to the two familiar examples of the euclidean and the elliptic spaces. Namely, *either the space is simply connected and all geodesics are congruent to a euclidean straight line, or the space Σ is not simply connected and all geodesics are congruent to one euclidean circle* (with the length of the shorter arc as distance). *Then Σ possesses a two-sheeted universal covering space Σ_1 which also satisfies the Axioms* A, B, C, D *and which shares the following property with an ordinary sphere. All geodesics of Σ_1 are congruent to one euclidean circle, and the geodesics passing through a given point X_1 of Σ_1 all pass through the other point X_1' of Σ_1 which lies over the same point of Σ as X_1.*

This theorem will be applied to obtain improvements on the theory of spaces with convex spheres as developed in chap. 4 of the author's *Metric methods in Finsler spaces and in the Foundations of geometry*(1). Finally the two-dimensional case will be discussed.

1. **Sphere like spaces.** The axioms which guarantee the existence of geodesics in the space Σ are these:

A. *Σ is metric.*

The distance of the two points A, B will be denoted by AB. We use the notation (ABC) to express that the three points A, B, C are different and that $AB+BC=AC$.

B. *Σ is finitely compact, that is, any bounded sequence of points has an accumulation point.*

Presented to the Society, September 13, 1943; received by the editors September 27, 1942.

(1) Annals of Mathematics Studies, no. 8, Princeton, 1942. This book will be referred to as B. It contains the references to the literature, which will therefore be omitted here. But it should be stated explicitly that in the special case of *flat spaces our main result was established long ago by Hamel* (see Hamel [2] in the bibliography of B). Moreover, it has been shown already in B that if the geodesic through any two distinct points is unique, each *single* geodesic is congruent to a euclidean circle or to a euclidean straight line.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

C. *Σ is convex, that is, for $A \neq C$, there is a point B with (ABC).*

D. *Every point P of Σ has a neighborhood ν with this property: For any two points $A \neq B$ in ν and a given $\epsilon > 0$ there is a positive $\delta(A, B) \leqq \epsilon$ to which there corresponds a unique point B_δ with*

$$(ABB_\delta) \text{ and } BB_\delta = \delta.$$

We shall now compile those consequences of Axioms A, B, C, D which will be needed in the present note. For the proofs the reader is referred to B.

A point set τ is called a *segment* if it is congruent to an interval $[a, b]$ of the real axis, that is, there is a mapping $t \to P(t)$ of $[a, b]$ on τ such that $P(t_1)P(t_2) = |t_1 - t_2|$ for any t_i in $[a, b]$. $P(t)$ is called an isometric representation of τ. The points $P(a)$ and $P(b)$ are the end points of τ. Any two points are the end points of at least one segment. $\mathfrak{t}(PQ)$ will denote a segment from P to Q. The following trivial property of segments will be used frequently.

(1) *If (ABC) and σ, τ are segments $\mathfrak{t}(AB)$ and $\mathfrak{t}(BC)$ respectively, then $\sigma + \tau$ is a segment $\mathfrak{t}(AC)$.*

When the spherical neighborhood $\nu(P, \rho)$ of P (consisting of those points X for which $PX < \rho$) satisfies Axiom D and $0 < \bar{\rho} \leqq \rho/6$ then $\nu(P, \bar{\rho})$ is a so-called *standard neighborhood* of P, that is, it has the following properties:

(2) *For any two distinct points A, B in $\nu(P, \bar{\rho})$ and a given $0 < t < 1$ there is exactly one point C with (ACB) and $AC = t \cdot AB$. And for any given $0 < s \leqq 2\bar{\rho}$ there is exactly one point D with (ABD) and $BD = s$.*

A point set g is called a *geodesic* if it is a map $t \to P(t)$ of the real axis $-\infty < t < \infty$ such that for every real t_0 a positive $\epsilon(t_0)$ exists for which the subarc $|t - t_0| \leqq \epsilon(t_0)$ of $P(t)$ represents a segment isometrically. $P(t)$ is called an isometric representation of g.

(3) *For a given segment $\mathfrak{t}(PQ)$ with an isometric representation $P'(t)$, $a \leqq t \leqq b$, $a < b$, there exists one and only one geodesic g which contains $\mathfrak{t}(PQ)$, and g admits an isometric representation $P(t)$ such that $P(t) = P'(t)$ for $a \leqq t \leqq b$.*

The simplest geodesics are the *straight lines*. They are by definition congruent either to a euclidean straight line or to a euclidean circle. More explicitly, if $P(t)$ represents a straight line isometrically then for any real t_1, t_2 either

(4a) $P(t_1)P(t_2) = |t_1 - t_2|$ (*open straight line*)

or an $\alpha > 0$ exists such that

(4b) $P(t_1)P(t_2) = \min |t_1 - t_2 + \nu\alpha|$, $\nu = 0, \pm 1, \pm 2, \cdots$ (*closed straight line*).

In the latter case g is congruent to a euclidean circle of radius $\alpha/2\pi$. Hence α is called the length $\lambda(g)$ of g. For any point A of g there is exactly one point A' on g (*the conjugate point to A on g*) for which g contains two segments $\mathfrak{t}(AA')$.

The point A' may also be characterized as the point on g for which no point Y with $(AA'Y)$ exists. For, if $A' \neq \bar{A} \subset g$ then $(A\bar{A}Y)$ with $Y = A'$. If a point Y with $(AA'Y)$ existed, it would obviously not lie on g. On the other

hand, it follows from (1) that a segment τ from A to A' on g forms together with any segment $\mathfrak{t}(A'Y)$ a segment τ' from A to Y. Let g' be the geodesic that contains τ'. Then $g' \neq g$ because $Y \not\subset g$. But (3) and the fact that τ is contained both in g and g' would yield $g = g'$.

A straight line contains with any two points P, Q also a segment $\mathfrak{t}(PQ)$. The converse holds too, so that we may state

(5) *A geodesic is a straight line if, and only if, with any two points P, Q it also contains a segment $\mathfrak{t}(PQ)$.*

In this form the theorem is not in B. However the proof of Theorem 3, chap. 1, §2, in B can easily be modified to yield this result.

As a consequence of (5) we have

(6) *If there is only one geodesic through any two distinct points, then all geodesics are straight lines.*

Namely, if g is any geodesic and A, B two different points on g, then a geodesic g' exists containing a given segment $\mathfrak{t}(AB)$ (compare (3)). Since, by hypothesis, $g = g'$, it follows that g contains a segment $\mathfrak{t}(AB)$, and hence must be a straight line by virtue of (5).

A space which satisfies the hypothesis of (6) is called an *S.L. space*, in particular an *open S.L. space* if all straight lines are open, a *closed S.L. space* if they are all closed.

As stated in the introduction, the main purpose of this note is to show that all S.L. spaces are either open or closed, and that in the latter case all straight lines have the same length.

In any space with Axioms A to D the straight lines have this property:

(7) *If two straight lines g and h have two different common points A and A' then either $g = h$ or else g and h are closed, have the same length, and A and A' are conjugate on both g and h.*

For if g is open or closed and A' not conjugate to A, then g contains a point Y with $(AA'Y)$ and a segment σ from A' to Y. The geodesic h contains a segment τ from A to A'. By virtue of (1) $\tau + \sigma$ is a segment $\mathfrak{t}(AY)$. Using (3) twice we find that the geodesic which contains $\tau + \sigma$ must coincide with h because it contains τ and with g because it contains σ. Hence, if $g \neq h$, the points A and A' are conjugate on both lines, so that $\lambda(g) = \lambda(h) = 2AA'$.

Let the space Σ satisfy Axioms A, B, C, D. If all geodesics of Σ are closed straight lines and if all geodesics which pass through a given point A pass through a fixed point A', the conjugate point to A in Σ, we call Σ *sphere like.* (The word spherical will be reserved for spaces congruent to a euclidean sphere.)

(8) *Let Σ be a sphere like space. Then all geodesics have the same length. The conjugate point A to A' in Σ is the conjugate point to A on every geodesic through A. The mapping $A \to A'$ of Σ on itself is an isometry. The geodesic through A and B is unique, when A and B are not conjugate.*

The first two assertions follow from (7).

There is a geodesic g through A and B. Any geodesic $g' \neq g$ through A intersects g, besides at A, only at A', hence g is the only geodesic through A and B. If we denote generally the conjugate point to X by X', the isometry of the mapping $A \to A'$ may be written as

(9) $XY = X'Y'$.

For $Y = X$ or $Y = X'$ the relation (9) is obvious. When $Y \neq X, X'$ the points X and Y are not conjugate. Hence the geodesic through X and Y is unique and contains both X' and Y'. Since $XX' = YY'$ and g is a closed straight line of length $2XX'$ we see that $XY = X'Y'$.

We show next that every sphere like space is covering space of a closed S.L. space.

THEOREM 1. *The non-ordered pairs $\overline{X} = (X, X')$ of a sphere like space Σ form with the metric*

(10) $\overline{XY} = \min(XY, XY')$

a closed S.L. space $\overline{\Sigma}$ in which all straight lines have length XX'.

Proof. Obviously $\overline{XX} = 0$. If $\overline{X} \neq \overline{Y}$, then $X \neq Y$ and $X \neq Y'$, hence $\overline{XY} > 0$. Moreover (9) yields

$$\overline{XY} = \min(XY, XY') = \min(YX, YX') = \overline{YX}.$$

The triangle inequality

(11) $\overline{XY} + \overline{YZ} \geqq \overline{XZ}$

follows from (9), (10) and these 4 inequalities:

$$XY + YZ \geqq XZ, \quad XY' + YZ = X'Y + YZ \geqq X'Z = XZ',$$

$$XY + YZ' \geqq XZ', \quad XY' + YZ' = X'Y + YZ' \geqq X'Z' = XZ,$$

which also show that the equality sign holds in (11) only if X, Y, Z are on one geodesic in Σ.

We see that $\overline{\Sigma}$ satisfies Axiom A. Clearly, Axiom B holds too. The metrization (10) applied to the points X of Σ which belong to a fixed geodesic g (the point X' lies then also on g) amounts to identifying diametrically opposite points on a circle, and yields therefore a closed straight line in $\overline{\Sigma}$ with length XX'. Hence any two distinct points $\overline{X}$, $\overline{Y}$ in $\overline{\Sigma}$ are on at least one closed straight line. This line contains a segment from X to Y, consequently $\overline{\Sigma}$ satisfies Axiom C.

For $XY < XX'/2$ we have

(12) $\overline{XY} = \min(XY, XY') = \min(XY, YY' - XY) = XY$.

Any two points X, Y of a spherical neighborhood $\nu(P, \rho)$ in Σ have distance $XY \leqq XP + PY < 2\rho$. Therefore it follows from (12) that the mapping $X \to \overline{X}$ maps the neighborhood $\nu(P, XX'/4)$ of P in Σ isometrically on the neighborhood $\nu(\overline{P}, XX'/4)$ in $\overline{\Sigma}$. Hence Axiom D holds in $\overline{\Sigma}$.

The characteristic property of a geodesic to be locally a segment shows together with the local isometry of the mapping $X \to \overline{X}$ that each geodesic $\bar{g}$ of $\overline{\Sigma}$ consists of those points $\overline{X} = (X, X')$ for which X belongs to a fixed geodesic in Σ. We have seen already that these points form a closed straight line of length XX'. Since the geodesic through two not conjugate points X and Y in Σ is unique, it follows that the geodesic through $\overline{X}$ and $\overline{Y}$ in $\overline{\Sigma}$ is unique, as soon as $\overline{X} \neq \overline{Y}$.

By means of Theorem 1 results on closed S.L. spaces lead to results on sphere like spaces. For instance Theorems 7 and 8 in B, chap. 5, §1, yield the following:

A sphere like space is spherical (that is, congruent to a euclidean sphere) *if it can be reflected in each of its straight lines.*

A symmetric sphere like space is spherical.

2. **The main theorem.** It is easy to see that *an open S.L. space is simply connected.* The segment $t(XY)$ connecting two arbitrary points X and Y is unique and depends continuously on the end points X and Y. Let $X(t)$, $a \leqq t \leqq b$, be any continuous curve and P any point. We denote by $X_\alpha(t)$, $0 \leqq \alpha \leqq 1$, the unique point on $t(PX(t))$ for which $PX_\alpha(t) = \alpha PX(t)$. The point $X_\alpha(t)$ depends continuously on both t and α, moreover $X_1(t) = X(t)$, $X_0(t) = P$. Hence varying α from 1 to 0 we obtain a deformation of $X(t)$ into P.

Similarly we see that *a sphere like space Σ is simply connected.* For let $X(t)$, $a \leqq t \leqq b$, be any continuous curve which does not cover all of Σ, and P' any point not on $X(t)$; finally let P be the conjugate point to P'. Then the segment $t(PX(t))$ is unique for every $X(t)$ and we may proceed as before.

We are going to show that an S.L. space $\overline{\Sigma}$ which is not open is not simply connected. This may be accomplished either by showing that a closed straight line cannot be contracted to a point, or by constructing a covering space Σ different from $\overline{\Sigma}$. We shall use the second method because it will yield directly a two-sheeted sphere like space Σ. Since on the one hand 2 is the minimum number of sheets and on the other hand a sphere like space is simply connected, we see that Σ is, topologically, the only covering space of $\overline{\Sigma}$. The original space is obtained from Σ as in Theorem 1, which implies that all geodesics in $\overline{\Sigma}$ are closed straight lines of the same length.

Let the S.L. space $\overline{\Sigma}$ contain at least one closed geodesic $\bar{g}$, and let $\overline{A}$ be an arbitrary point on $\bar{g}$. The unique geodesic through the two different points $\overline{X}$, $\overline{Y}$ of $\overline{\Sigma}$ will be denoted by $g(\overline{XY})$. When $\overline{D} \neq \overline{A}$ and $g(\overline{AD})$ is closed, then any geodesic $g(\overline{AX})$ with $\overline{X} \subset \nu(\overline{D}, \epsilon)$ is closed provided $\epsilon > 0$ is sufficiently small (compare B, chap. 3, §1, Theorem 3).

Call $\bar{\beta}$ the locus of the conjugate points to $\overline{A}$ on the different closed geodesics through $\overline{A}$. Finally let $\nu(\overline{A}, \rho)$ be a standard neighborhood of $\overline{A}$ (compare (2)) and $\bar{\kappa}$ its boundary, that is, the locus $\overline{XA} = \rho$.

We are going to define a mapping $\overline{X} \to X^*$ of $\overline{\Sigma} - \bar{\beta}$ on $\nu(\overline{A}, \rho)$. We put $A^* = \overline{A}$. If $\overline{A} \neq \overline{X} \subset \overline{\Sigma} - \bar{\beta}$, the segment $t(\overline{AX})$ is unique, hence there is exactly

one point X^* with $(AX^*\overline{X})$ and

$$A^*X^* = \rho(\overline{AX}/(1+\overline{AX}))((2+\lambda g(\overline{AX}))/\lambda g(\overline{AX})),$$

where $\lambda g(\overline{AX})$ is the length of the geodesic $g(\overline{AX})$ and

$$(2+\lambda g(\overline{AX}))/\lambda g(\overline{AX}) = 1 \text{ if } \lambda g(\overline{AX}) = \infty.$$

Because the length of a geodesic depends continuously on the geodesic (see B, chap. 3, §1) the mapping $\overline{X}\to X^*$ is one-to-one and continuous. Call κ^* the subset of those points X^* of $\bar{\kappa}$ for which $g(\overline{A}X^*)$ is closed. We map the point X^* of κ^* on the conjugate point $\overline{X}$ to $\overline{A}$ on $g(\overline{A}X^*)$. We then obtain a continuous mapping α of $\Sigma^*=\nu(\overline{A},\rho)+\kappa^*$ on all of $\bar{\Sigma}$. Each point $\overline{X}$ of $\bar{\beta}$ has two images X_1^* and X_2^*, which are diametrically opposite points of $\bar{\kappa}$.

We now assign a definite neighborhood $u(X^*)$ to every point of Σ^* as follows: If $\overline{X}$ is not on $\bar{\beta}$ we let $u(\overline{X})$ be a standard neighborhood of $\overline{X}$ which does not meet $\bar{\beta}$, and $u(X^*)$ the image $u(\overline{X})\alpha^{-1}$ of $u(\overline{X})$. If $\overline{X}\subset\bar{\beta}$ and X_1^* and X_2^* are its two images we choose a positive $\eta<\rho/2$ so that $g(\overline{AZ})$ is closed for $\overline{Z}\subset\nu(X_1^*,\eta)$ or $\overline{Z}\subset\nu(X_2^*,\eta)$. We choose a standard neighborhood $u(\overline{X})$ of $\overline{X}$ such that $g(\overline{AY})$ meets $\nu(X_1^*,\eta)$ and $\nu(X_2^*,\eta)$ when $\overline{Y}\subset u(\overline{X})$. (This is possible because $g(\overline{AX}_\nu)\to g(\overline{AX})$ when $\overline{X}_\nu\to\overline{X}$.) Let $\bar{u}_i$, $i=1, 2$, be the set of those points $\overline{Y}$ in $u(\overline{X})$ for which a point $\overline{Z}_i$ in $\nu(X_i^*,\eta)$ with $(\overline{Y}\overline{Z}_i\overline{A})$ exists. We have $\bar{u}_1+\bar{u}_2=u(\overline{X})$, and $\bar{u}_1$ and $\bar{u}_2$ have only the points $\overline{Y}$ of $\bar{\beta}\cdot u(\overline{X})$ in common. For if $(\overline{Y}\overline{Z}_i\overline{A})$ and $\overline{Z}_i\subset\nu(X_i^*,\eta)$, $i=1, 2$, then

$$\overline{Z}_1\overline{Z}_2 > X_1^*X_2^* - 2\eta = 2\rho - 2\eta > \rho > 2\eta.$$

If $\overline{Y}$ were not conjugate to $\overline{A}$, the segment $t(\overline{AY})$ would be unique and contain both $\overline{Z}_1$ and $\overline{Z}_2$. But then

$$\overline{Z}_1\overline{Z}_2 = |\,\overline{AZ}_1 - \overline{AZ}_2\,| < (\rho+\eta) - (\rho-\eta) = 2\eta.$$

The image $u(X_i^*)$ of $\bar{u}_i$ we define as neighborhood of X_i^*.

Now let Σ^{**} be congruent to Σ^* and designate by X^{**} the image of X^* under a definite congruent mapping μ of Σ^* on Σ^{**}. Let X_1^* and X_2^* be any two diametrically opposite points of κ^*. We then identify X_1^* with X_2^{**} and X_1^{**} with X_2^*. The sum of Σ^* and Σ^{**} will be called Σ. The mappings α and $\mu\alpha$ define together a continuous mapping α of Σ on $\bar{\Sigma}$. Under α^{-1} every point $\overline{X}$ of $\bar{\Sigma}$ has exactly two images X^* and X^{**} in Σ. We say that X^* and X^{**} lie over $\overline{X}$.

With every point of Σ we also associate a neighborhood. Namely if $\overline{X}\not\subset\bar{\beta}$ then we take the previously defined $u(X^*)$ as neighborhood of X^* and $u(X^{**})=u(X^*)\mu$ as neighborhood of X^{**}. If $\overline{X}\subset\bar{\beta}$ we take $u(X_1^*)\mu+u(X_2^*)$ as neighborhood of $X_1^{**}=X_2^*$ and $u(X_1^*)+u(X_2^*)\mu$ as neighborhood of $X_1^*=X_2^{**}$. The neighborhoods $u(X^*)$ and $u(X^{**})$ of the two points X^* and X^{**} over $\overline{X}$ are disjoint and α maps both $u(X^*)$ and $u(X^{**})$ homeomorphically on the standard neighborhood $u(\overline{X})$ of $\overline{X}$.

We metrize Σ in the following manner: For a continuous curve c with a representation $P(t)$, $0 \leqq t \leqq 1$, in Σ we define the number $\lambda(c)$ as the length of the image $\overline{P}(t) = P(t)\alpha$ in $\overline{\Sigma}$ and put for any two points X, Y in Σ

(13) $XY = \inf \lambda(c_{XY})$,

where c_{XY} traverses all continuous curves from X to Y. XY is always defined and finite. For, any point of X^* of Σ^* (or X^{**} of Σ^{**}) can be connected to A^* (or A^{**}) by a path whose image is one of the possibly two segments from $\overline{X}$ to $\overline{A}$, and the points A^* and A^{**} can be connected by a curve whose image is a given closed straight line through $\overline{A}$ in $\overline{\Sigma}$.

The relations $XX = 0$, $XY = YX$, $XY + YZ \geqq XZ$ are obvious and so is $XY > 0$, for $\overline{X} \neq \overline{Y}$ (we put generally $Z\alpha = \overline{Z}$). If $\overline{X} = \overline{Y}$ the relation $XY > 0$ follows from $u(X^*) \cdot u(X^{**}) = 0$.

We show next that with this metrization of Σ the mapping α is locally isometric. We see from the definition of $\lambda(c)$ and XY that

(14) $XY \geqq \overline{X}\overline{Y}$.

If $u(\overline{X}) = v(\overline{X}, \delta)$ we put $v(\overline{X}) = v(\overline{X}, \delta/2)$. Under the homeomorphic mapping α of $u(X)$ on $u(\overline{X})$ where X stands for a definite one of the two images of $\overline{X}$, the set $v(X)$ may correspond to $v(\overline{X})$. Let $A, B \subset v(X)$. Since $\overline{A}, \overline{B} \subset v(\overline{X})$ the (unique) segment $t(\overline{A}\overline{B})$ lies completely in $u(\overline{X})$ so that the image of $t(\overline{A}\overline{B})$ in $u(X)$ has the same length as $t(\overline{A}\overline{B})$, that is $\overline{A}\overline{B}$. Hence it follows from (13) and (14) that

(15) $AB = \overline{A}\overline{B}$ for $A, B \subset v(X)$.

We conclude from (15) that $v(X)$ satisfies the requirements of D for the point X; since X was an arbitrary point of Σ we see that Axiom D holds.

Because of the local isometry (15) the number $\lambda(c)$ equals the length of c as curve in Σ.

Axiom B now follows readily. For if $PX_\nu < d$ then $\overline{P}\overline{X}_\nu < d$ because of (14). Hence $\{\overline{X}_\nu\}$ contains a subsequence $\{\overline{X}_{\nu_n}\}$ converging to a point $\overline{X}$ in $\overline{\Sigma}$. (15) implies that with proper notations the images $\{X_{\nu_n}^*\}$ and $\{X_{\nu_n}^{**}\}$ of $\{\overline{X}_{\nu_n}\}$ tend to the images X^* and X^{**} of $\overline{X}$ respectively. The sequence $\{X_{\nu_n}\}$ contains infinitely many points of at least one of the two sequences $\{X_{\nu_n}^*\}$ and $\{X_{\nu_n}^{**}\}$, which proves Axiom B.

To prove Axiom C we observe that because of B and (13) a curve c_{XZ} exists which connects two given points X and Z and such that

$$XZ = \lambda(c_{XZ})$$

(compare Property Le in B, chap. 1, §1). The additivity of the arclength and the triangle inequality yield in the well known fashion that $XY + YZ = XZ$ for every point of c_{XZ}.

We conclude from the local isometry (15) of Σ and $\overline{\Sigma}$ that the image $\bar{g} = g\alpha$ of any geodesic in Σ is a geodesic in $\overline{\Sigma}$, and that at least one geodesic g in Σ lies over every geodesic $\bar{g}$ in $\overline{\Sigma}$.

Moreover, if g and h both lie over $\bar{g}$ and have a common point P, then

$g=h$, because g and h coincide in a neighborhood of P (compare (3)). It follows that there is *only one* geodesic which lies over a given geodesic $\bar{g}$ of $\overline{\Sigma}$. For if g_1 and g_2 lie over $\bar{g}$ we select a point G_i on g_i such that $\overline{G}_1 \neq \overline{G}_2$. There is a geodesic k through G_1 and G_2. Its image $\bar{k}$ intersects $\bar{g}$ at both $\overline{G}_1$ and $\overline{G}_2$, hence $\bar{g}=\bar{k}$, and it follows from $k \cdot g_i \supset G_i$ that $k=g_i$ so that also $g_1=g_2$. We may now conclude

(16) *If a geodesic g of Σ contains X^*, then g also contains X^{**}.*

For g lies over a geodesic $\bar{g}$ through $\overline{X}$, and there is a geodesic g_1 over $\bar{g}$ through X^{**}. Hence $g=g_1 \supset X^{**}$. Before we apply this result to the present problem we notice a consequence which will be important in the theory of the convexity of spheres.

THEOREM 2. *In a closed S.L. space each geodesic $\bar{g}$ intersects the locus $\beta(\overline{A})$ of the conjugate points to an arbitrary point $\overline{A}$ (possibly $\bar{g} \subset \beta(\overline{A})$).*

Returning to our spaces Σ and $\overline{\Sigma}$ we see that the geodesics through $\overline{A}$ must all be closed, for an open geodesic would have two images, one in the interior of Σ^* and one in Σ^{**}.

Since the length $\lambda(\bar{g})$ of a geodesic g depends continuously on $\bar{g}$, and all geodesics through $\overline{A}$ are closed, the lengths of the geodesics through $\overline{A}$ are bounded. Hence the space $\overline{\Sigma}$ is bounded and therefore compact. We show next

(17) *Each geodesic g in Σ is a closed straight line.*

Since $\overline{\Sigma}$ is compact Σ is compact, hence open straight lines cannot exist. Because of (5) it is sufficient to show that g contains with any two points X, Y also a segment $t(XY)$. Let $\overline{X} \neq \overline{Y}$. Consider a segment $t(XY)$ and a geodesic h which contains this segment $t(XY)$. Then $\bar{g}$ and $\bar{h}$ both contain $\overline{X}$ and $\overline{Y}$, hence $\bar{g}=\bar{h}$ and $g=h$, so that g contains $t(XY)$. Now let $\overline{X}=\overline{Y}$, and consider a sequence of points $Y_\nu \neq Y$ on g which tend to Y. For large ν we have $\overline{Y}_\nu \neq \overline{X}$. Hence g contains a segment $t(XY_\nu)$. A suitable subsequence of $\{t(XY_\nu)\}$ tends to a segment $t(XY)$. As the image of $\bar{g}$ the geodesic g is closed as a point set, hence $t(XY) \subset g$, which proves (17).

(16) and (17) show that Σ is a sphere like space. Moreover $\overline{\Sigma}$ can be obtained from Σ as in Theorem 1. This implies that all geodesics in $\overline{\Sigma}$ have length X^*X^{**}. We formulate our results as follows:

THEOREM 3. *Let the space $\overline{\Sigma}$ satisfy Axioms* A, B, C, D *and let there be only one geodesic through two distinct points of $\overline{\Sigma}$.*

Then $\overline{\Sigma}$ is either simply connected and all geodesics are congruent to euclidean straight lines.

Or $\overline{\Sigma}$ has a sphere like space Σ as two-sheeted universal covering space, from which $\overline{\Sigma}$ is obtained by identifying conjugate points. The geodesics of $\overline{\Sigma}$ are all congruent to euclidean circles and have the same length.

3. **Convexity of spheres.** A sphere $K(A, \rho)$ is the locus of those points

X for which $XA=\rho$. The set $\nu(A, \rho)$ is by definition the interior of $K(A, \rho)$. When α is the common length of the straight lines of a closed S.L. space $K(A, \alpha/2)$ is the locus $\beta(A)$ of the conjugate points to A, and $K(A, \rho)$ is empty for $\rho>\alpha/2$. In the following the radius ρ of a sphere is always assumed to be not greater than $\alpha/2$.

Theorems 2 and 3 enable us to simplify and to complete the theory of spaces with convex spheres as developed in B, chap. 4. We shall indicate the new method here.

In an open S.L. space we would naturally say that $K(A, \rho)$ is convex if a segment $t(XY)$ whose end points X, Y are in or on $K(A, \rho)$ lies except for X, Y in the interior of $K(A, \rho)$. But this definition cannot be applied to closed S.L. spaces, since then not even all spheres of an elliptic space would be convex. In euclidean geometry an equivalent definition of convexity is available in terms of supporting planes. One is therefore led to ask whether convexity can be expressed satisfactorily in terms of the existence of supporting geodesics.

We call the straight line t of a sphere like space or an S.L. space a *tangent* of $K(A, \rho)$ at Z if points $X_\nu \neq Y_\nu$ on $K(A, \rho)$ exist which tend to Z and such that the line $g(X_\nu Y_\nu)$ tends to t. We say that $K(A, \rho)$ *is convex if no tangent of* $K(A, \rho)$ *contains points of* $\nu(A, \rho)$. The spheres of an elliptic space are convex in this sense.

On the other hand one might consider dropping the distinction between the interior and exterior of a sphere and simply require that the spheres are surfaces of order 2. More exactly we say that a set μ of a sphere like space or of an S.L. space has *order not greater than* n if any straight line intersects μ in at most n different points, or else is contained in μ, and that μ has *order* n if it has order *not greater than* n *but not order less than or equal to* $n-1$. We can then show the following:

(18) *Let the sphere* $K(A, \rho)$ *of an arbitrary S.L. space* Σ *not be the locus* $\beta(A)$ *of the conjugate points to* A *and have order* 2. *Then* $K(A, \rho)$ *is convex.*

Assume for an indirect proof that a point Z on $K(A, \rho)$ and a tangent t of $K(A, \rho)$ at Z, which contains a point P of $\nu(A, \rho)$, exist. By definition t is limit of a sequence of straight lines $g(X_\nu Y_\nu)$ where $X_\nu \neq Y_\nu$, $X_\nu A = Y_\nu A = \rho$ and $X_\nu \to Z$, $Y_\nu \to Z$. Choose P_ν on $g(X_\nu Y_\nu)$ such that $P_\nu \to P$.

First let Σ be an open S.L. space. Then either of the subrays of $g(X_\nu Y_\nu)$ with origin P_ν contains a point of $K(A, \rho)$ (because PT varies continuously as T traverses $g(X_\nu Y_\nu)$). The ray which contains X_ν will also contain Y_ν, at least for large ν, because $X_\nu Y_\nu \to 0$. But then $g(X_\nu Y_\nu)$ would intersect $K(A, \rho)$ in at least three points.

Next let Σ be closed and $K(A, \rho) \neq \beta(A)$. Because of Theorem 2 the line $g(X_\nu Y_\nu)$ contains a point Q_ν of $\beta(A)$. Then $Q_\nu A > \rho$. For the same reason as before either subarc of $g(X_\nu Y_\nu)$ from P_ν to Q_ν contains a point of $K(A, \rho)$,

and for large ν the arc which contains X_ν also contains Y_ν, so that $\mathfrak{g}(X_\nu Y_\nu)$ would intersect $K(A, \rho)$ in at least three different points. This result will be completed by:

(19) *If the spheres $K(A, \rho)$ of a closed S.L. space have order not greater than 2 then $\beta(A)$ has order 1.*

We have to show that $\beta(A)$ contains the line g if it contains the two distinct points X and Y of g. Call δ_1 and δ_2 the two subarcs of g from X to Y. If the theorem were not true, at least one of the arcs, δ_1 say, would contain a point P with $PA=\rho-\epsilon_1$, $\epsilon_1>0$. Then δ_2 cannot contain a point Q with $QA=\rho-\epsilon_2$, $\epsilon_2>0$, because the sphere $K(A, \rho-\min(\epsilon_1, \epsilon_2)/2)$ would contain at least four points of $\mathfrak{k}$ without containing g. Hence $K(A, \alpha/2)=\beta(A)$ would have to contain δ_2 without containing g, and would not have order less than or equal to 2.

The results (18) and (19) lead to the following theorem:

THEOREM 4. *In any S.L. space the following properties are equivalent:*

I. *The spheres are convex.*

II. *The spheres have order not greater than 2.*

III. *A given point P has exactly one foot on a given straight line g; or else PX is constant for $X \subset g$.*

IV. *If the point P does not lie on the line g, then either $PX=$const. for $X \subset g$ and all lines $\mathfrak{g}(PX)$ are perpendicular[2] to g, or else there is exactly one perpendicular through P to g.*

A proof of Theorem 4 may be obtained by combining our present results with the results in B, chap. 4, §§1 and 3. Although many simplifications are now possible we refrain from giving a complete proof here in order to avoid duplication.

B, chap. 4, §1, Theorem 3, states that I implies III, and B, chap. 4, §3, Theorem 1 contains the equivalence of II and III. Our present relations (18) and (19) show that II implies I. Finally B, chap. 4, §3, Theorems 4, 5 and 6 yield together with the present Theorems 2 and 3 that IV follows from II. Therefore we have to show here only that IV implies III. Let P not be conjugate to all points of g and not on g (otherwise III is trivial) and consider the perpendicular $\mathfrak{g}(PF)$, $F \subset g$, through P to g. Since not all points of g are conjugate to P, F (as a foot of P) is not conjugate to P, so that $\mathfrak{g}(PF)$ contains a point Q with (QPF). According to the definition of a perpendicular F is a foot of Q on g, therefore we have for every $X \neq F$ on g

$$PX > QX - QP \geqq QF - QP = PF,$$

which means that F is the only foot of P on g.

---

[2] The straight line $h \neq g$ is a perpendicular to g if h intersects g and all points of h have this intersection as foot on g.

Theorems 1 and 4 yield, together with B, chap. 4, §2, Theorem 1, the following characterization of the elliptic spaces.

THEOREM 5a. *A closed S.L. space of dimension not less than 3, which has one of the properties* I, II, III, *or* IV, *is congruent to a finite dimensional elliptic space.*

A sphere $K(A, \rho)$ of a sphere like space of dimension not less than 2 whose geodesics have length 2α is not convex as soon as $\rho > \alpha/2$. But the point sets which carry $K(A, \rho)$ and $K(A', \alpha-\rho)$ coincide and one of these spheres is convex. Therefore we conclude from Theorems 5a and 1:

THEOREM 5b. *Let Σ be a sphere like space of dimension not less than 3, whose geodesics have length 2α. If the spheres of radius $\rho \leqq \alpha/2$ are convex, or if one of the properties* II, III, IV *holds, Σ is congruent to a euclidean sphere.*

In B, chap. 4, §2 an example of Hamel was quoted to show that a closed S.L. space is not even then necessarily elliptic when its geodesics are the straight lines of a projective space. By means of Theorem 1 we see from this example that there are metrizations of a euclidean sphere (of any dimension not less than 2) in which the geodesics are the euclidean great circles but the metric is not spherical. The condition of Theorem 5b may therefore not be omitted.

4. **Two-dimensional spaces.** The characterization (5) of the straight lines suggests the problem to determine all spaces which satisfy Axioms A, B, C, D and in which the geodesics are straight lines. As solution one would conjecture the S.L. spaces and the sphere like spaces. The proof seems difficult for higher dimensions but it is easy for two.

THEOREM 6. *Let each geodesic of a two-dimensional space Σ (with Axioms* A, B, C, D*) contain a segment $\mathrm{t}(XY)$ when it contains X and Y. Then Σ is either homeomorphic to the euclidean plane and an open S.L. space, or homeomorphic to the projective plane and a closed S.L. space, or homeomorphic to a 2-sphere and sphere like.*

Proof. If no two geodesics have more than one common point, the assertion is contained in B, chap. 3, §2, Theorem 1. Let therefore two different geodesics g and h which have the two common points A and A' exist. It follows from (7) that they have no further common points and that g and h are closed with A and A' as conjugate points on both lines.

Call σ_1, σ_2 and τ_1, τ_2 the two segments $\mathrm{t}(AA')$ on g and h respectively. Because of B, chap. 1, §4, Theorem 4, a standard neighborhood $\nu(A, \rho)$ of A is homeomorphic to a circular disk. Let S_i, T_j be the point on σ_i and τ_j respectively with $S_iA = T_jA = \rho$, and call g_X a geodesic which carries a segment $\mathrm{t}(A'X)$, where X is an interior point of the unique segment $\mathrm{t}(S_iT_j)$. By virtue of B, chap. 1, §4, Theorem 3, g_X intersects $\mathrm{t}(S_iA)+\mathrm{t}(AT_j)$. The point A is

the only possible intersection because each intersection of g_X with g or h other than A' is conjugate to A' on g_X and g or h.

Therefore the (unique) segment $t(AX)$ lies on g_X. The set $\sum_x t(AX)$ with $X \subset \sum_{i,j} t(S_iT_j)$ covers a whole neighborhood of A, hence all geodesics through A are closed straight lines and pass through A'. The geodesic $\mathfrak{g}(AX)$ is unique because X is not conjugate to A, and varies continuously with X. Hence the space is homeomorphic to the 2-sphere.

Consider now any point $B \neq A, A'$. There is a geodesic k through A and B. Let l be any other geodesic through B. Since l crosses k at B and is a closed Jordan curve it follows from the topological structure of a sphere that l intersects k at some other point B'. We see as before that every geodesic through B' passes also through B. Since B was arbitrary the space is sphere like.

In an elliptic space we have complete duality between point and hyperplane. The dual of a given theorem is obtained by interchanging a point A with the hyperplane $\beta(A)$ formed by the conjugate points to A. In a general closed S.L. space of dimension $n \geqq 3$, hyperplanes are not defined and questions of duality become void.

But we may ask whether there are *duality relations which hold in every closed S.L. plane.* $\beta(A)$ will in general not be a straight line. But in the elliptic case $\beta(A)$ may be characterized as the straight line which has maximal distance from A. We shall show the following theorem:

THEOREM 7. *For every point $\overline{A}$ of a closed S.L. plane $\overline{\Sigma}$ there is exactly one straight line $g(\overline{A})$ which has maximal distance from $\overline{A}$. The correspondence $\overline{A} \to g(\overline{A})$ maps $\overline{\Sigma}$ topologically on the set π of all geodesics in $\overline{\Sigma}$.*

By Theorem 3 this theorem is equivalent to the following statement on a two-dimensional sphere like space Σ.

For a given point A in Σ there is exactly one straight line $g(A)$ which has maximal distance from A. When A' is conjugate to A, then $g(A) = g(A')$. The mapping $(A, A') \to g(A)$ of the pairs of conjugate points on the geodesics in Σ is topological.

We shall prove this fact. Because of Theorem 6 the space Σ is homeomorphic to the 2-sphere. Designate by 4α the common length of the geodesics in Σ and by $d(A, g)$ the distance of the point A from the geodesic g. If A is not conjugate to X then the line $\mathfrak{g}(AX)$ is unique for $X \subset g$ and intersects g also in X'. We see from

(20) $2\alpha = XA + AX' = X'A' + A'X$

that

(21) $d(A, g) = d(A', g)$

and $\min(AX, AX') \leqq \alpha$, hence

(22) $d(A, g) \leqq \alpha$. *The equality sign holds if and only if $AX \equiv \alpha$ for $X \subset g$.*

The number $d(A, g)$ depends continuously on A and g. For a given point A there is a line g_0 for which $d(A, g)$ reaches its maximum $m \leqq \alpha$.

(23) $d(A, g_0)=m=\alpha$ if and only if A has two feet on g_0 which are conjugate to each other.

For if F and F' are feet of A on g_0, we see from (20) that $d(A, g_0)=AF=AF'=\alpha$. Conversely if $d(A, g_0)=\alpha$, then $AX=AX'=\alpha$ for any X on g_0 because of (22).

The main point of the proof is

(24) *If g_0 has maximal distance from A, then any open segment on g_0 of length 2α contains a foot of A.*

Assume for an indirect proof that (24) is not true. Then it follows from (23) that g_0 contains no two conjugate feet. Since the feet of A form a closed subset of g_0, there is a closed segment s of length 2α on g_0 which contains no foot of A, so that $AX>m+\epsilon$ for $X\subset s$ and a suitable $\epsilon>0$.

The endpoints Y, Y' of s are conjugate. Call s' the other segment $\mathfrak{t}(YY')$ on g_0. A foot F of A is an interior point of s'. Since $AF=m<\alpha$, there are points Z with (AFZ). The line $g_Z=\mathfrak{g}(YZ)$ is unique and passes through Y'. Call s_Z' the subsegment $\mathfrak{t}(YY')$ of g_Z which contains Z and s_Z the other subsegment $\mathfrak{t}(YY')$ of g_Z. For $Z\to F$ we have $g_Z\to g_0$, $s_Z\to s$, $s_Z'\to s'$ and $d(A, g_Z)\to d(A, g_0)$. Consequently, when FZ is sufficiently small each foot F_Z of A on g_Z belongs to s_Z' so that $d(A, s_Z')=d(A, g_Z)$. But s' separates A from s_Z' in the domain containing A and bounded by s and s_Z'. Hence $d(A, g_Z)=d(A, s_Z')>d(A, s')=d(A, g_0)$, which contradicts the definition of g_0.

(25) Any line $h\neq g_0$ contains an open segment of length 2α which is free from feet of A.

This statement is trivial for $A\subset h$. Let therefore A not be on h and designate by D and D' the intersections of g_0 and h, by μ the side of g_0 which contains A, and by t the segment $\mathfrak{t}(DD')$ on h which is not contained in μ. For any interior point G of t the segment $\mathfrak{t}(AG)$ intersects g at some point $H\neq D, D'$. Hence G is no foot of A on h, because otherwise

$$d(A, h) = AG > AH \geqq d(A, g_0),$$

which contradicts the definition of g_0. We see that the interior of s satisfies (25).

(24) and (25) show that g_0 is the only line which has maximal distance from A. We designate this line henceforth by $g(A)$. Moreover $g(A)=g(A')$.

(26) If $B\neq A, A'$ then $g(A)$ contains an open segment of length 2α which is free from feet of B.

The statement is obvious for $B\subset g(A)$. If B does not lie on $g(A)$, let it lie on the same side of $g(A)$ as A. The line $\mathfrak{g}(AB)$ contains a point H on $g(A)$ with (ABH). If H is a foot of A on $g(A)$, then any $X\neq H$ on g satisfies the inequality

$$BX > AX - AB \geqq AG - AB = BG,$$

so that G is the only foot of B on $g(A)$ and (26) holds. If H is not a foot of A,

we traverse $g(A)$ from H on, in either direction, until we meet the first foot of A. We may thus obtain the feet F_1, F_2. Because of (24) we have $F_1F_2 < 2\alpha$. To prove (26) it suffices to see that $t(F_1F_2)$ contains all feet of B. If G is any point on $g(A) - t(F_1F_2)$, the segment $t(AG)$ intersects $t(AF_1) + t(AF_2)$ at some point R. Let $R \subset t(AF_1)$. Since $F_1F_2 < 2\alpha$ the point F_1 is different from the conjugate point H' to H, hence we obtain, using that F_1 is a foot of A,

$$BF_1 < BR + RF_1 = BR + AF_1 - AR \leqq BR + AG - AR \leqq BR + GR = BG.$$

This shows that G is not a foot of B.

We conclude from (24) and (26) that $g(A) \neq g(B)$ for $A \neq B$; therefore the mapping $A \to g(A)$ is one-to-one. The continuity follows immediately from the continuity of $d(A, g)$ in A and g.

It is easily seen that all geodesics of Σ form a set π homeomorphic to the projective plane. Let μ map π topologically on Σ. The mapping $A \to g(A) \to g(A)\mu$ maps Σ on the subset Σ' of those points which correspond to geodesics which have maximal distance from some point of Σ. Since the mapping of Σ on Σ' is topological, we have $\Sigma = \Sigma'$. Hence $g(A)$ traverses all of π as A traverses Σ.

This theorem throws some light on the difficulties of the problem to determine the closed S.L. planes with convex circles, because it shows how weak an assumption the convexity of the circles is in two dimensions.

ILLINOIS INSTITUTE OF TECHNOLOGY,
CHICAGO, ILL.

SPECTRAL THEORY. I

CONVERGENCE TO PROJECTIONS

BY

NELSON DUNFORD

Spectral theory has for its origin the classical canonical reduction of a matrix, and it is this classical theory that provides the pattern for the developments presented in this paper. The finite dimensional linear vector space of the classical theory is here replaced by a complex Banach space $\mathfrak{X}$ and the matrix in that theory by a continuous linear operator T in the space $\mathfrak{X}$. In this situation there are always three and sometimes four types of convergence, namely, the uniform, the strong, the weak, and, in case $\mathfrak{X}$ consists of measurable functions, convergence almost everywhere. While our primary aim is to give necessary and sufficient conditions for convergence (in each of the four types) of a sequence $Q_n(T)$ of polynomials in T to a specified kind of limit operator, we have, and not merely incidentally but of necessity as well as for its own sake, developed a parallel to many of the algebraic aspects of classical matrix theory. In fact the chief tools in the convergence theory here presented are theorems corresponding to (a) the minimal equation theorem for matrices, (b) the theorem of Sylvester concerning the determinant of a polynomial in a matrix, and (c) the various decompositions of the whole space into a direct sum of its subspaces determined by the spectrum of a matrix. This analogue to classical matrix algebra plays its role almost exclusively in the questions involving uniform convergence (which of the four types seems to be the most difficult one to handle; especially in making the proofs of necessity rather than those of sufficiency) and the reader interested only in the other three types of convergence will find the last section fairly complete in itself as far as these questions are concerned.

The type of limit operator $E=\lim_n Q_n(T)$ with which we shall be chiefly concerned is a projection of the whole space $\mathfrak{X}$ on the manifold $\mathfrak{M}[P]$ consisting of all $x\in\mathfrak{X}$ where, for a specified polynomial P, we have $P(T)x=0$. In seeking conditions for uniform convergence of a given sequence $Q_n(T)$ towards such a limit, one is faced with two fairly obvious necessary conditions, one of which is $P(T)Q_n(T)\to 0$. A third condition which will make the set of three both necessary and sufficient may be any one of three types, namely, (a) of *spectral type* and asserting that the roots of $P(\lambda)=0$ be either in the resolvent set of T or else poles of the resolvent, or (b) of *algebraic type* and asserting that the manifold $\mathfrak{N}[P]=P(T)\mathfrak{X}$ is a closed complement of $\mathfrak{M}[P]$, or (c) of *topological type* and asserting that for every root λ of $P(\lambda)=0$ of

Presented to the Society, April 24, 1943; received by the editors October 6, 1942, and, in revised form, March 26, 1943.

multiplicity ν we have $(\lambda I-T)^{\nu+1}\mathfrak{X}$ closed. One way of visualizing the difference between strong convergence and uniform is through the corresponding algebraic condition for strong convergence which is $\overline{\mathfrak{N}}[P]\oplus\mathfrak{M}[P]=\mathfrak{X}$.

The ergodic theorem falls within the scope of this theory, for in the ergodic theorem we are dealing with the case of a projection on the fixed points of an operator, that is $P(T)=I-T$. As the general situation shows, the conditions for convergence of the particular sequence $n^{-1}\sum_{m=0}^{n-1}T^m$ of polynomials are, if worded properly, equally valid for any sequence of polynomials $Q_n(T)$ with $Q_n(1)\to 1$, but for the particular sequence $n^{-1}\sum_{m=0}^{n-1}T^m$ the condition $P(T)Q_n(T)=(I-T)n^{-1}\sum_{m=0}^{n-1}T^m=(I-T^n)/n\to 0$ takes the relatively simple form of $T^n/n\to 0$. It is the condition $T^n/n\to 0$ and not the customary one of $|T^n|\leqq M$ which is essential in the ergodic theory.

Since an element of a normed ring (in the sense of Gelfand) when considered as a linear operator on that ring has its norm as an operator the same as its norm in the ring, the purely algebraic results of §2 as well as those results in §3 concerning uniform convergence may be thought of as results concerning an element T of a normed ring.

The indigenous connection between the spectrum of an operator T and questions of convergence of sequences of functions $Q_n(T)$ of T is most readily seen (but not completely comprehended) by an examination of the spectrum of a finite matrix. For this reason and also because the finite case furnishes some insight into the puzzles of the infinite one we shall first present certain aspects of classical matrix theory in a form roughly as they shall appear in the general case.

1. **Spectral theory in a finite dimensional space.** We shall not discuss this case as completely as we shall the infinite one but enough will be given to indicate the general trend of ideas. In fact we shall merely derive a formula (equivalent to the classical canonical reduction) which will serve as a basis for an operational calculus and show how questions of convergence may be settled by spectral considerations. Except for the first two definitions and lemmas which are not concerned with the dimension or the topology in $\mathfrak{X}$ (and which will be needed in the infinite case) we are concerned here with a linear vector space $\mathfrak{X}$ of dimension $n>0$ over the field of complex numbers (or any algebraically closed field) and a linear operator T (that is an $n\times n$ matrix) which maps $\mathfrak{X}$ into all or part of itself.

1.1. DEFINITION. *Let P be a polynomial, λ a complex number, and ν a positive integer or 0. We define*

$$\mathfrak{N}[P]=P(T)\mathfrak{X}, \qquad \mathfrak{M}[P]=\mathop{\mathcal{E}}_{x\in\mathfrak{X}}[P(T)x=0],$$

$$\mathfrak{N}_\lambda^\nu=(\lambda I-T)^\nu\mathfrak{X}, \qquad \mathfrak{M}_\lambda^\nu=\mathop{\mathcal{E}}_{x\in\mathfrak{X}}[(\lambda I-T)^\nu x=0],$$

where I is the identity operator in $\mathfrak{X}$.

1.2. DEFINITION. *If* $\mathfrak{M}, \mathfrak{M}_1, \cdots, \mathfrak{M}_k$ *are linear manifolds in* $\mathfrak{X}$ *we shall write* $\mathfrak{M}=\mathfrak{M}_1\oplus\cdots\oplus\mathfrak{M}_k$ *if and only if* $\mathfrak{M}_i\subset\mathfrak{M}$, $i=1, \cdots, k$, *and every* $x\in\mathfrak{M}$ *has a unique representation* $x=x_1+\cdots+x_k$ *with* $x_i\in\mathfrak{M}_i$, $i=1, \cdots, k$. *We shall write* $\mathfrak{M}=\mathfrak{M}_1+\cdots+\mathfrak{M}_k$ *if and only if* $\mathfrak{M}_i\subset\mathfrak{M}$, $i=1, \cdots, k$, *and every* $x\in\mathfrak{M}$ *can be represented as* $x=x_1+\cdots+x_k$ *with* $x_i\in\mathfrak{M}_i$, $i=1, \cdots, k$.

1.3. LEMMA. *If* P, Q *are polynomials with no common roots then*

$$\mathfrak{M}[P\cdot Q] = \mathfrak{M}[P]\oplus\mathfrak{M}[Q].$$

Clearly $\mathfrak{M}[P], \mathfrak{M}[Q]\subset\mathfrak{M}[P\cdot Q]$. Since P and Q have no common factors there are polynomials R, S with

$$R(\lambda)P(\lambda)+S(\lambda)Q(\lambda)\equiv 1,$$
$$R(T)P(T)+S(T)Q(T) = I,$$
$$R(T)P^2(T)+S(T)P(T)Q(T) = P(T).$$

Thus if $P(T)Q(T)x=0$, that is, if $x\in\mathfrak{M}[P\cdot Q]$, then $P(T)(x-y)=0$ where $y=R(T)P(T)x\in\mathfrak{M}[Q]$. This decomposition is unique, for the second of the above identities shows that $u=0$ if $u\in\mathfrak{M}[P]\cdot\mathfrak{M}[Q]$.

1.4. LEMMA. *Let* $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ *be a polynomial whose distinct roots are* $\lambda_1, \cdots, \lambda_k$. *Let* $P_i(\lambda)=P(\lambda)/(\lambda_i-\lambda)^{\nu_i}$. *Then*

$$\mathfrak{M}[P] = \mathfrak{M}_{\lambda_1}^{\nu_1}\oplus\cdots\oplus\mathfrak{M}_{\lambda_k}^{\nu_k}$$

and

$$\mathfrak{X} = \mathfrak{N}[P_1]+\cdots+\mathfrak{N}[P_k].$$

The first conclusion is proved by a repeated application of Lemma 1.1 and the second by observing that since $P_1(\lambda), \cdots, P_k(\lambda)$ have no common factor there are polynomials $R_1, \cdots, R_k$ with

$$R_1(\lambda)P_1(\lambda)+\cdots+R_k(\lambda)P_k(\lambda)\equiv 1,$$
$$R_1(T)P_1(T)+\cdots+R_k(T)P_k(T) = I.$$

1.5. DEFINITION. *The index* ν *of a complex number* λ *is the smallest positive integer or zero such that* $(T-\lambda I)^{\nu+1}x=0$ *implies* $(T-\lambda I)^{\nu}x=0$.

Thus the index ν of λ is positive if and only if λ is a characteristic number of the matrix T. In what follows we use the symbols $\lambda_1, \cdots, \lambda_k$ for the distinct characteristic numbers of T and $\nu_1, \cdots, \nu_k$ for their indices.

1.6. LEMMA. *If* $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ *then* $P(T)=0$ *and hence*

$$\mathfrak{X} = \mathfrak{M}_{\lambda_1}^{\nu_1}\oplus\cdots\oplus\mathfrak{M}_{\lambda_k}^{\nu_k}.$$

To see that $P(T)=0$ let x be an arbitrary vector in $\mathfrak{X}$. Since the dimension

of $\mathfrak{X}$ is n there are constants $\alpha_0, \cdots, \alpha_n$ with $\alpha_0 x+\alpha_1 Tx+\cdots+\alpha_n T^n x=0$, that is, $Q(T)x=0$ for some polynomial Q. If Q is written in factored form then $\prod(\mu_i I-T)^{m_i}x=0$ and it is clear that this equation remains valid if all factors with μ_i not a characteristic number are removed, and from the definition of ν_i we see that the other factors may be changed to have the desired exponents. Thus $P(T)x=0$, $P(T)=0$. The second conclusion follows from Lemma 1.4.

1.7. DEFINITION. *Let E_{λ_i}, $i=1, \cdots, k$, be the projections which, in view of Lemma* 1.6, *exist and satisfy $I=E_{\lambda_1}+\cdots+E_{\lambda_k}$, $E_{\lambda_i}^2=E_{\lambda_i}$, $E_{\lambda_i}\mathfrak{X}=\mathfrak{M}_{\lambda_i}^{\nu_i}$, $E_{\lambda_i}E_{\lambda_j}=0$, $i\neq j$.*

Note that since $T\mathfrak{M}_{\lambda_i}^{\nu_i}\subset\mathfrak{M}_{\lambda_i}^{\nu_i}$ we have $E_{\lambda_i}Tx=E_{\lambda_i}\sum_j TE_{\lambda_j}x=TE_{\lambda_i}x$, that is $TE_{\lambda_i}=E_{\lambda_i}T$.

1.8. DEFINITION. *Let $\mathcal{F}(T)$ be the class of all complex functions of a complex variable which are regular at every point of an open set (not necessarily connected) containing the characteristic numbers $\lambda_1, \cdots, \lambda_k$ of T. (The open set upon which $f\in\mathcal{F}(T)$ is regular may vary with f.) For $f\in\mathcal{F}(T)$ we define*

$$(*)\qquad f(T)=\sum_{i=1}^{k}\sum_{m=0}^{\nu_i-1}\frac{(T-\lambda_i I)^m}{m!}f^{(m)}(\lambda_i)E_{\lambda_i}.$$

1.9. THEOREM. *If $f, g\in\mathcal{F}(T)$ and α, β are complex numbers then*
(a) *$\alpha f+\beta g\in\mathcal{F}(T)$ and $(\alpha f+\beta g)(T)=\alpha f(T)+\beta g(T)$,*
(b) *$f\cdot g\in\mathcal{F}(T)$ and $(f\cdot g)(T)=f(T)\cdot g(T)$,*
(c) *if $f(\lambda)=\sum_{n=0}^{m}\alpha_n\lambda^n$ then $f(T)=\sum_{n=0}^{m}\alpha_n T^n$.*

Statements (a) and the first part of (b) are obvious. To prove that $(f\cdot g)(T)=f(T)\cdot g(T)$ let us write, for $h\in\mathcal{F}(T)$,

$$D_{\lambda_i}(h, T)=\sum_{m=0}^{\nu_i-1}\frac{(T-\lambda_i I)^m}{m!}h^{(m)}(\lambda_i).$$

Then since $(T-\lambda_i I)^q E_{\lambda_i}=0$ for $q\geqq\nu_i$ we have

$$\begin{aligned}D_{\lambda_i}(f, T)D_{\lambda_i}(g, T)E_{\lambda_i}&=\sum_{p=0}^{2\nu_i-2}\sum_{n=0}^{p}\frac{(T-\lambda_i I)^p}{n!(p-n)!}f^{(n)}(\lambda_i)g^{(p-n)}(\lambda_i)E_{\lambda_i}\\&=\sum_{p=0}^{2\nu_i-2}\frac{(T-\lambda_i I)^p}{p!}\left[\sum_{n=0}^{p}{}_pC_n f^{(n)}(\lambda_i)\cdot g^{(p-n)}(\lambda_i)\right]E_{\lambda_i}\\&=\sum_{p=0}^{\nu_i-1}\frac{(T-\lambda_i I)^p}{p!}(f\cdot g)^{(p)}(\lambda_i)E_{\lambda_i}\\&=D_{\lambda_i}(f\cdot g, T).\end{aligned}$$

Hence

$$f(T)\cdot g(T) = \left[\sum_{i=1}^{k} D_{\lambda_i}(f, T)E_{\lambda_i}\right]\cdot\left[\sum_{i=1}^{k} D_{\lambda_i}(g, T)E_{\lambda_i}\right]$$

$$= \sum_{i=1}^{k} D_{\lambda_i}(f, T)D_{\lambda_i}(g, T)E_{\lambda_i}$$

$$= \sum_{i=1}^{k} D_{\lambda_i}(f\cdot g, T)E_{\lambda_i} = (f\cdot g)(T),$$

which proves (b). In view of (a) and (b) statement (c) need only be verified for the special cases of $f(\lambda)=1$, $f(\lambda)=\lambda$ and (c) for these polynomials is an immediate consequence of Definition 1.7.

1.10. THEOREM. *If $f, g\in\mathcal{F}(T)$ then $f(T)=g(T)$ if and only if $f^{(n)}(\lambda_i)=g^{(n)}(\lambda_i)$, $i=1, \cdots, k$; $n=0, \cdots, \nu_i-1$. Thus for a polynomial P we have $P(T)=0$ if and only if $P(\lambda)$ contains $\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ as a factor.*

The sufficiency of the condition is obvious from (*) and its necessity follows from the definition of ν_i as follows. In view of part (a) of Theorem 1.9 we may and shall assume that $f(T)=0$. Let $x\neq 0$, $(T-\lambda_i I)x=0$ so that by (*) and 1.7 we have $0=f(T)x=f(\lambda_i)x$, $f(\lambda_i)=0$. Next, in case $\nu_i>1$, pick x so that $(T-\lambda_i I)^2x=0$, $(T-\lambda_i I)x\neq 0$ then as before $0=f(T)x=f(\lambda_i)x+(T-\lambda_i I)f^{(1)}(\lambda_i)x=f^{(1)}(\lambda_i)(T-\lambda_i I)x$, and so $f^{(1)}(\lambda_i)=0$. A repetition of this argument clearly yields $f^{(n)}(\lambda_i)=0$, $n=0, 1, \cdots, \nu_i-1$.

1.11. THEOREM. *Let $f_n\in\mathcal{F}(T)$, then $f_n(T)$ converges if and only if $f_n^{(j)}(\lambda_i)$ converges for each $i=1, \cdots, k$, and $j=0, \cdots, \nu_i-1$.*

The sufficiency of the conditions is clear from (*) while their necessity follows from the argument used in the proof of the preceding theorem.

The next theorem is proved in the same way.

1.12. THEOREM. *Let $f, f_n\in\mathcal{F}(T)$ then $f_n(T)\to f(T)$ if and only if* $\lim f_n^{(j)}(\lambda_i)=f^{(j)}(\lambda_i)$, $i=1, \cdots, k; j=0, \cdots, \nu_i-1$.

1.13. *An illustration.* If $f_n(\lambda)=n^{-1}\sum_{m=0}^{n-1}\lambda^m$ we have $f_n(\lambda)$ converging if and only if $|\lambda|\leqq 1$, while $f^{(j)}(\lambda)$ converges for all $j=1, 2, \cdots$ providing $|\lambda|<1$ but the first derivative $f^{(1)}(\lambda)$ fails to converge if $|\lambda|=1$. Hence (1.11) shows that $n^{-1}\sum_{m=0}^{n-1}T^m$ converges if and only if all roots λ_i of T have $|\lambda_i|\leqq 1$ and if in addition $\nu_i=1$ for every root λ_i with $|\lambda_i|=1$. Next consider the statement $T^n/n\to 0$. An application of (1.12) to the functions $f_n(\lambda)=\lambda^n/n$ shows us that $T^n/n\to 0$ if and only if the roots λ_i of T satisfy the same conditions required in the preceding example. *Hence $n^{-1}\sum_{m=0}^{n-1}T^m$ converges if and only if $T^n/n\to 0$.* It is not pure coincidence that these two statements concerning convergence have precisely the same spectral interpretation. It is a consequence of the fact that the statement $T^n/n\to 0$ is equivalent to $(I-T)n^{-1}\sum_{m=0}^{n-1}T^m$

$=(I-T^n)/n\to 0$ and is a special case of a general principle which will be developed more completely in §3.

1.14. *Concluding remarks.* The formula

$$\text{(i)} \qquad f(T)=\sum_{i=1}^{k}\sum_{m=0}^{\nu_i-1}\frac{(T-\lambda_i I)^m}{m!}f^{(m)}(\lambda_i)E_{\lambda_i}$$

of 1.8 becomes

$$\text{(ii)} \qquad f(T)=\sum_{i=1}^{k}f(\lambda_i)E_{\lambda_i}$$

for a matrix T whose characteristic numbers λ_i all have index 1. In particular (ii) holds if T is an Hermitian matrix. This case has been generalized in an extensive literature to the concept of a self-adjoint operator in Hilbert space where (ii) takes the well known form of $f(T)=\int_{-\infty}^{+\infty}f(\lambda)dE_\lambda$, and in recent years has been given, by M. H. Stone(1) and I. Gelfand(2), an algebraic topological form. One difference between the ring of polynomials in a general matrix T and the ring of polynomials in an Hermitian matrix T is that (as formulas (i) and (ii) show) the former ring is isomorphic with a direct sum of reduced polynomial rings while the latter is isomorphic with a direct sum of identical fields, that is, isomorphic with a linear space of scalar functions.

If the projections E_{λ_i} in (i) are eliminated from (i) by using the Hermite interpolation formula then (i) takes the very natural form

$$\text{(iii)} \qquad f(T)=\frac{1}{2\pi i}\int_C \frac{f(\lambda)}{\lambda I-T}\,d\lambda,$$

where C is the boundary of a set of small circles containing $\lambda_1, \cdots, \lambda_k$. If C_i is a small circle about λ_i then (iii) gives

$$\text{(iv)} \qquad E_{\lambda_i}=\frac{1}{2\pi i}\int_{C_i}\frac{d\lambda}{\lambda I-T}.$$

Formula (iii) was suggested by E. Cartan(3) as a possible basis for an operational calculus for infinite matrices and was used (only in an incidental way) by Gelfand. Formulas (iii) and (iv) were discovered by Fantappié(4) who was working with finite matrices, and recently (iv) has been extended to the infinite case by Lorch(5), who proved that an operator E defined by the equa-

(1) M. H. Stone, *A general theory of spectra.* I, Proc. Nat. Acad. Sci. U. S. A. vol. 26 (1940) pp. 280–283.

(2) I. Gelfand, *Normierte Ringe*, Rec. Math. (Mat. Sbornik) N. S. vol. 51 (1941) pp. 3–24.

(3) The suggestion was made in a letter to G. Giorgi. For this reference as well as for others pertaining to an operational calculus see C. C. MacDuffee, *The theory of matrices*, Berlin, 1933.

(4) L. Fantappiè, *La calcul des matrices*, C. R. Acad. Sci. Paris vol. 186 (1928) pp. 619–621.

(5) E. R. Lorch, *The spectrum of linear transformations*, Trans. Amer. Math. Soc. vol. 52 (1942) pp. 238–248.

tion $E=(2\pi i)^{-1}\int_C(\lambda I-T)^{-1}d\lambda$ is a projection, and that the algebra of such projections is isomorphic with the Boolean algebra of all open and closed subsets of the spectrum of T.

2. **An operational calculus and some algebraic properties of linear operators.** The procedure of §2 used to discuss the general linear operator T differs only slightly from that of §1. We first define the class $\mathcal{F}(T)$ and the operator $f(T)$ for $f\in\mathcal{F}(T)$ and then define the projections corresponding to the E_{λ_i} of §1. The fundamental algebraic theorems (Theorems 2.12 through 2.29) may then be formulated in terms of these concepts.

2.1. *Preliminary concepts and definitions.* From now on the symbol T will stand for a bounded linear operator on a complex Banach space $\mathfrak{X}$. By the *resolvent set* $\rho(T)$ will be meant the set of all complex numbers λ such that $(\lambda I-T)^{-1}$ exists as an everywhere defined (and hence continuous) linear operator. For $\lambda\in\rho(T)$ we shall write $R_\lambda(T)$ for $(\lambda I-T)^{-1}$. The set $\rho(T)$ is an open set containing all λ with $|\lambda|>|T|$. By the *spectrum* $\sigma(T)$ of T is meant all complex numbers $\lambda\notin\rho(T)$. The spectrum $\sigma(T)$ is a closed point set contained in the circle $|\lambda|\leqq|T|$. If the space $\mathfrak{X}$ contains more than one point (and this we assume) the spectrum $\sigma(T)$ is not empty (see Taylor(6)) and we write $|\sigma(T)|$ for l.u.b. $|\lambda|$ where λ varies over $\sigma(T)$. The *point spectrum* $\sigma_p(T)$ consists of all λ such that $(\lambda I-T)^{-1}$ does not exist, the *continuous spectrum* $\sigma_c(T)$ of those λ for which $(\lambda I-T)^{-1}$ exists as an unbounded operator with domain dense in $\mathfrak{X}$, and the *residual spectrum* $\sigma_r(T)$ of those λ for which $(\lambda I-T)^{-1}$ exists with domain nondense in $\mathfrak{X}$. It is readily shown that if $(\lambda I-T)^{-1}$ exists as a bounded operator with domain dense in $\mathfrak{X}$ then $\lambda\in\rho(T)$, so that the point sets $\rho(T)$, $\sigma_p(T)$, $\sigma_c(T)$, $\sigma_r(T)$ are disjoint and their union is the whole complex plane. The function $R_\lambda(T)$ satisfies the *resolvent equation* $R_\lambda(T)-R_\mu(T)=(\mu-\lambda)R_\mu(T)R_\lambda(T)$, λ, $\mu\in\rho(T)$, and is *regular*, that is, $dR_\lambda(T)/d\lambda$ exists, at every point $\lambda\in\rho(T)$. If $f(\lambda)$ is any vector (or operator) valued function, regular and single-valued at every point of a closed domain D whose boundary C is a finite number of rectifiable Jordan curves then (1) $\int_C f(\lambda)d\lambda=0$, (2) $f(\lambda)=(2\pi i)^{-1}\int_C f(\xi)(\xi-\lambda)^{-1}d\xi$, for λ in the interior of D, (3) $f(\lambda)$ has derivatives of all orders at every point of D, (4) $f^{(n)}(\lambda)=(2\pi i)^{-1}n!\int_C f(\xi)(\xi-\lambda)^{-n-1}d\xi$, for λ in the interior of D, (5) the Taylor expansion $\sum_{n=0}^{\infty}(\lambda-\xi)^n f^{(n)}(\xi)/n!$ converges uniformly to $f(\lambda)$ for λ in any circle $|\lambda-\xi|\leqq r$ contained in the interior of D. If $f(\lambda)$ is regular and single-valued at each point of an annulus enclosed by two concentric circles C_1 and C_2 with center λ_0 and if C is any circle between C_1 and C_2, then for every λ in this annulus

$$f(\lambda)=\sum_{n=-\infty}^{+\infty}a_n(\lambda-\lambda_0)^n,\quad a_n=\frac{1}{2\pi i}\int_C\frac{f(\lambda)d\lambda}{(\lambda-\lambda_0)^{n+1}},\qquad n=0,\pm 1,\pm 2,\cdots.$$

(6) A. E. Taylor, *The resolvent of a closed transformation*, Bull. Amer. Math. Soc. vol. 44 (1938) pp. 70–74.

The above statements concerning regular functions may be proved as follows. For a linear functional over $\mathfrak{X}$, that is a point $x^* \in \mathfrak{X}^*$, we have $x^* \int_C f(\lambda)d\lambda = \int_C x^* f(\lambda)d\lambda$ and thus by Cauchy's theorem $x^* \int_C f(\lambda)d\lambda = 0$. Hence by the Hahn-Banach extension theorem (which was proved for complex linear functionals over a complex Banach space by Bohnenblust and Sobczyk([7])) we have $\int_C f(\lambda)d\lambda = 0$. This proves (1) and (2) is proved in the same way while the rest of the statements follow in the usual manner.

Many standard theorems in the theory of linear operators on a Banach space are readily seen to hold in a complex Banach space and these will be used when needed. In particular the various corollaries of the Hahn-Banach extension theorem including the *principle of uniform boundedness* will be used. Also the *interior mapping theorem* or, more specifically, its corollary which asserts that *a closed everywhere defined linear operator is continuous* will be used. This shows that the projections E, $E' = I - E$ determined by *closed* complementary linear manifolds in $\mathfrak{X}$ are continuous.

An operational calculus

2.2. Definition. *By a T-admissible domain is meant an open set D of complex numbers having the following properties*:

(1) *D is a finite sum of connected open sets D_i with $\overline{D}_i \overline{D}_j = 0$, $i \neq j$.*

(2) *The boundary C of D consists of a finite number of disjoint closed rectifiable Jordan curves contained in the resolvent set $\rho(T)$ of T. We shall sometimes write $D = D(T)$ to mean that D is a T-admissible domain.*

2.3. Definition. *The class $\mathcal{F}(T)$ is the class of all complex functions each of which is regular and single-valued at each point of the closure of some $D = D(T)$ which contains the spectrum $\sigma(T)$ of T.*

2.4. Lemma. *Let D, D^* be T-admissible domains with $\sigma(T) \subset D^* \subset \overline{D}^* \subset D$. Let $\phi(\lambda)$ be regular and single-valued on $\overline{D}$. Then*

$$\int_C \phi(\lambda) R_\lambda(T) d\lambda = \int_{C^*} \phi(\lambda) R_\lambda(T) d\lambda,$$

where C, C^* are the boundaries of D, D^* respectively.

The domain D is made up of a finite sum of disjoint sets and taking the integral around C is the same as taking the integral around each connected component and adding. Thus, without loss of generality, we may and shall confine our attention to the case where D is connected. Suppose first that D is simply connected and that D^* is connected. Since $\sigma(T) \subset D^*$ none of the inside boundaries of D^* can surround a part of $\sigma(T)$ and hence the integral around every inside boundary of D^* vanishes. Thus the integral $\int_{C^*}$ reduces

([7]) H. F. Bohnenblust and A. Sobczyk, *Extensions of functionals on complex linear spaces*, Bull. Amer. Math. Soc. vol. 44 (1938) pp. 91–93.

to the integral around the outside boundary of D^* which may be joined by a cut to the boundary of D to prove the lemma in this case. Next suppose only that D is simply connected. This case reduces to the first case by making cuts joining up the various parts of D^*. Now suppose the lemma has been proved for the case where D is $(n-1)$-tuply connected. We now take D to be n-tuply connected. If at least one of the inside boundaries of D lies outside of all the outside boundaries of the various parts of D^* it may be joined by a cut to the outside boundary of D, thus reducing D to a $(n-1)$-tuply connected domain. If every inside boundary C_i of D lies in one of the holes of some part of D^* and is thus surrounded by some uniquely determined nearest part C_i^* of the boundary C^* then $\int_{C_i} = \int_{C_i^*}$. Also any inside boundary C_i^* of D^* which does not surround such a C_i nor any portion of D^* has the property that $\int_{C_i^*} = 0$ and this type of C_i^* may be dropped. These observations reduce the problem to the case where (i) every inside boundary C_i of C is contained in at least one of the holes in some part of D^* and (ii) every inside boundary of D^* surrounds either (α) an inside boundary of D or (β) one or more parts of D^*. Here parts (α) and (β) are not mutually exclusive. Now consider an inside boundary C_i^* of D^* which comes under (ii)(β). The various portions of D^* within C_i^* may be joined by cuts and one of them to C_i^*. This process reduces the problem to the case where every inside C_i is in one and only one inside C_i^* and every inside C_i^* contains one C_i. Now it is clear that the integral around the remaining (uncut) outside boundaries of D^* is equal to the integral around the outside boundary of D and the integral around such inside boundary of D^* is equal to the integral around the corresponding boundary of D and thus $\int_C = \int_{C^*}$.

2.5. LEMMA. *Let D, D^* be T-admissible domains with $\sigma(T) \subset DD^*$. Then there is a T-admissible domain D_1 with $\sigma(T) \subset D_1 \subset DD^*$. If ϕ is regular and single-valued at each point of $\overline{D}$ and at each point of $\overline{D}^*$ then*

$$\int_C \phi(\lambda) R_\lambda(T) d\lambda = \int_{C^*} \phi(\lambda) R_\lambda(T) d\lambda,$$

where C, C^* are the boundaries of D, D^* respectively.

The first part of this lemma follows from the Heine-Borel Theorem and the second part then follows from the preceding lemma.

2.6. DEFINITION. *Let $f \in \mathcal{F}(T)$, $D = D(T)$, $D \supset \sigma(T)$ and let f be regular and single-valued on $\overline{D}$. Let C be the boundary of D. We define*

$$f(T) = \frac{1}{2\pi i} \int_C f(\lambda) R_\lambda(T) d\lambda.$$

The preceding lemma shows that $f(T)$ is independent of D.

2.7. THEOREM. *If $f, g\in\mathcal{F}(T)$ then*
(1) *$f+g\in\mathcal{F}(T)$ and $f(T)+g(T)=(f+g)(T)$,*
(2) *$f\cdot g\in\mathcal{F}(T)$ and $f(T)\cdot g(T)=(f\cdot g)(T)$.*

Statement (1) as well as the first part of (2) is obvious. To see that $f(T)\cdot g(T)=(f\cdot g)(T)$ let C be the boundary of a domain $D=D(T)$ such that $\sigma(T)\subset D$ and such that f and g are both regular and single-valued on $\overline{D}$. Let C_j be the boundaries of the component parts of D. Then

$$f(T)=\frac{1}{2\pi i}\sum_{j=1}^{k}\int_{C_j}f(\lambda)R_\lambda(T)d\lambda,\qquad g(T)=\frac{1}{2\pi i}\sum_{j=1}^{k}\int_{C_j}g(\lambda)R_\lambda(T)d\lambda,$$

$$(*)\qquad -4\pi^2 f(T)\cdot g(T)=\sum_{n=1}^{k}\sum_{m=1}^{k}\left(\int_{C_n}f(\lambda)R_\lambda(T)d\lambda\right)\left(\int_{C_m}g(\lambda)R_\lambda(T)d\lambda\right).$$

Now

$$\left(\int_{C_n}f(\lambda)R_\lambda(T)d\lambda\right)\left(\int_{C_m}g(\mu)R_\mu(T)d\mu\right)=\int_{C_n}\int_{C_m}f(\lambda)g(\mu)R_\lambda(T)R_\mu(T)d\lambda d\mu$$

$$=-\int_{C_n}\int_{C_m}f(\lambda)g(\mu)\frac{R_\lambda(T)-R_\mu(T)}{\lambda-\mu}d\lambda d\mu$$

$$=-\int_{C_m}g(\mu)d\mu\left[\int_{C_n}\frac{f(\lambda)R_\lambda(T)}{\lambda-\mu}d\lambda-R_\mu(T)\int_{C_n}\frac{f(\lambda)}{\lambda-\mu}d\lambda\right].$$

Hence if $n\neq m$ and $\mu\in C_m$ we have $\int_{C_n}f(\lambda)(\lambda-\mu)^{-1}d\lambda=0$ and

$$\left(\int_{C_n}f(\lambda)R_\lambda(T)d\lambda\right)\left(\int_{C_m}g(\mu)R_\mu(T)d\mu\right)$$

$$=-\int_{C_m}g(\mu)d\mu\int_{C_n}\frac{f(\lambda)R_\lambda(T)}{\lambda-\mu}d\lambda$$

$$=-\int_{C_n}f(\lambda)R_\lambda(T)d\lambda\int_{C_m}\frac{g(\mu)}{\lambda-\mu}d\mu=0,$$

and thus (*) becomes

$$(**)\qquad -4\pi^2 f(T)\cdot g(T)=\sum_{n=1}^{k}\int_{C_n}f(\lambda)R_\lambda(T)d\lambda\int_{C_n}g(\mu)R_\mu(T)d\mu$$

$$=-\sum_{n=1}^{k}\int_{C_n}\int_{C_n}f(\lambda)g(\mu)\frac{R_\lambda(T)-R_\mu(T)}{\lambda-\mu}d\lambda d\mu.$$

Now using each point of C_n as a center construct a circle of radius so small that f, g are both regular in the closed circle and also such that the radius is less than half the distance from the point on C_n to the remaining contours

bounding D. A finite number of these circles cover the boundary C_n of the component D_n of D. Let $\mathcal{E}_n$ be the sum of the interiors of this finite number of circles and let $D_n^* = \mathcal{E}_n + D_n$. Let C_n^* be the boundary of D_n^*. Clearly we have $\sigma(T)D_n \subset D_n^*$, $\bar{D}_n \subset D_n^*$. Then

$$\int_{C_n} f(\lambda)\frac{R_\lambda(T) - R_\mu(T)}{\lambda - \mu}\,d\lambda = \int_{C_n^*} f(\lambda)\frac{R_\lambda(T) - R_\mu(T)}{\lambda - \mu}\,d\lambda$$

$$= \int_{C_n^*} \frac{f(\lambda)R_\lambda(T)\,d\lambda}{\lambda - \mu} - 2\pi i f(\mu)R_\mu(T)$$

and (**) may be written as

$$4\pi^2 f(T)\cdot g(T) = \sum_{n=1}^{k}\Big[2\pi i\int_{C_n} f(\mu)g(\mu)R_\mu(T)\,d\mu$$

$$- \int_{C_n^*} f(\lambda)R_\lambda(T)\Big(\int_{C_n}\frac{g(\mu)\,d\mu}{\lambda - \mu}\Big)\,d\lambda\Big].$$

For $\lambda \in C_n^*$ we have $\int_{C_n} g(\mu)(\lambda - \mu)^{-1}\,d\mu = 0$ and so

$$f(T)\cdot g(T) = \frac{1}{2\pi i}\sum_{n=1}^{k}\int_{C_n} f(\mu)g(\mu)R_\mu(T)\,d\mu$$

$$= \frac{1}{2\pi i}\int_C f(\mu)g(\mu)R_\mu(T)\,d\mu = (f\cdot g)(T).$$

2.8. THEOREM. *If $f(\lambda)$ is a polynomial in λ or, more generally, if $f(\lambda) = \sum_{n=0}^{\infty}\alpha_n\lambda^n$ is a series convergent in a circle $|\lambda| \leqq \rho$ with $\rho > |T|$ then $f(T) = \sum_{n=0}^{\infty}\alpha_n T^n$.*

Let C be the circle $|\lambda| = \rho$ so that

$$f(T) = \frac{1}{2\pi i}\int_C f(\lambda)R_\lambda(T)\,d\lambda = \frac{1}{2\pi i}\int_C f(\lambda)\sum_{n=0}^{\infty} T^n\lambda^{-n-1}\,d\lambda$$

$$= \sum_{n=0}^{\infty} T^n\frac{1}{2\pi i}\int_{C_n}\frac{f(\lambda)}{\lambda^{n+1}}\,d\lambda = \sum_{n=0}^{\infty}\alpha_n T^n.$$

2.9. THEOREM. *If $f \in \mathcal{F}(T)$ then $f(\sigma(T)) = \sigma(f(T))$.*

Let $\lambda \in \sigma(T)$ and let $g(\xi)$ be defined for ξ in the domain of definition of f by the formula $g(\xi) = (f(\lambda) - f(\xi))(\lambda - \xi)^{-1}$ so that $g \in \mathcal{F}(T)$ and, by 2.7 and 2.8, we have $f(\lambda)I - f(T) = (\lambda I - T)g(T)$ which shows that $f(\lambda) \in \sigma(f(T))$. Conversely let $\mu \in \sigma(f(T))$ and suppose that $\mu \notin f(\sigma(T))$. Then the function $h(\xi) = (f(\xi) - \mu)^{-1}$ belongs to $\mathcal{F}(T)$ and, by 2.7 and 2.8, $h(T)(f(T) - \mu I) = I$, which contradicts the fact that $\mu \in \sigma(f(T))$.

2.10. THEOREM. *Let $f\in\mathcal{F}(T)$, $g\in\mathcal{F}(f(T))$ and let $F(\xi)=g(f(\xi))$. Then $F\in\mathcal{F}(T)$ and $F(T)=g(f(T))$.*

Suppose g is regular and single-valued on $\overline{D}^*$ where $\sigma(f(T))=f(\sigma(T))\subset D^*$ and let C^* be the boundary of D^*. Pick $D=D(T)$ such that $\sigma(T)\subset D$, f is regular and single-valued on $\overline{D}$ and $f(\overline{D})\subset D^*$. Then if C is the boundary of D we have from 2.7

$$R_\lambda(f(T)) = \frac{1}{2\pi i}\int_C \frac{R_\xi(T)}{\lambda - f(\xi)}\,d\xi, \qquad \lambda\in C^*,$$

and

$$\begin{aligned} g(f(T)) &= \frac{1}{2\pi i}\int_{C^*} g(\lambda)R_\lambda(f(T))\,d\lambda \\ &= \left(\frac{1}{2\pi i}\right)^2 \int_C\int_{C^*} \frac{g(\lambda)\,d\lambda}{\lambda - f(\xi)}\,R_\xi(T)\,d\xi \\ &= \frac{1}{2\pi i}\int_C g(f(\xi))R_\xi(T)\,d\xi = F(T). \end{aligned}$$

The elementary algebraic rules of operation given by 2.7, $\cdots$, 2.10 inclusive will be used from this point on without explicit reference to the theorem in question.

2.11. DEFINITION. *By a spectral set of T will be meant any subset σ of $\sigma(T)$ which is both open and closed in $\sigma(T)$. If σ is a spectral set of T the symbol σ' will be used for the complement of σ in $\sigma(T)$ so that σ' is also a spectral set of T.*

It is readily seen that a set σ of complex numbers is a spectral set of T if and only if it is of the form $\sigma=\sigma(T)D$ where $D=D(T)$, and that for any spectral set σ of T there are T-admissible domains D, D' with

$$\sigma = \sigma(T)D, \qquad \sigma' = \sigma(T)D', \qquad \overline{DD'} = 0.$$

Such domains are said to be *complementary domains* and D' will be called a *complement* of D. If C, C' are the boundaries of D, D' respectively and we define

$$E_\sigma[T] = \frac{1}{2\pi i}\int_C R_\lambda(T)\,d\lambda, \qquad E_{\sigma'}[T] = \frac{1}{2\pi i}\int_{C'} R_\lambda(T)\,d\lambda,$$

then it is seen from 2.7 and 2.8 that $E_\sigma[T]$, $E_{\sigma'}[T]$ are projections reducing T and that

$$I = E_\sigma[T] + E_{\sigma'}[T], \qquad E_\sigma[T]\cdot E_{\sigma'}[T] = 0.$$

With every spectral set σ of T is associated the closed linear manifold $\mathfrak{X}_\sigma[T]\subset\mathfrak{X}$ and defined as $\mathfrak{X}_\sigma[T]=E_\sigma[T]\mathfrak{X}$. Thus if $\sigma_1, \cdots, \sigma_k$ are disjoint spectral sets of T and $\sigma=\sigma_1+\cdots+\sigma_k$ we have

$$E_\sigma[T] = E_{\sigma_1}[T] + \cdots + E_{\sigma_k}[T], \qquad E_{\sigma_i}[T]\cdot E_{\sigma_j}[T] = 0, \qquad i \neq j,$$
$$\mathfrak{X}_\sigma[T] = \mathfrak{X}_{\sigma_1}[T] \oplus \cdots \oplus \mathfrak{X}_{\sigma_k}[T].$$

When there is no chance of confusion we shall sometimes write E_σ and $\mathfrak{X}_\sigma$ in place of $E_\sigma[T]$ and $\mathfrak{X}_\sigma[T]$ respectively. If the spectral set σ reduces to a single point λ we shall write E_λ, $\mathfrak{X}_\lambda$ for E_σ, $\mathfrak{X}_\sigma$ respectively.

2.12. THEOREM. *Let σ be a spectral set of T. Then when T is considered as an operator in the space $\mathfrak{X}_\sigma$ it has σ for its spectrum. Furthermore for $\lambda \in \sigma$ any one of the following statements is true if and only if it is valid when T is regarded as an operator in $\mathfrak{X}_\sigma$: λ is in the point spectrum of T, λ is in the residual spectrum of T, λ is in the continuous spectrum of T, λ is a pole of order ν for $R_\xi(T)$, λ is an essential singularity for $R_\xi(T)$, λ is an isolated point of the spectrum.*

Let $\lambda \notin \sigma$. We shall first show that $(\lambda I - T)$ is a 1-1 bicontinuous map of $\mathfrak{X}_\sigma$ into all of itself. Let $D = D(T)$ be such that $\sigma = \sigma(T)D$, $\lambda \notin \overline{D}$, and let D' be a complement to D. Define $f \in \mathcal{F}(T)$ so that $f(\xi) = (\lambda - \xi)^{-1}$ for ξ in a neighborhood of $\overline{D}$ and $f(\xi) = 0$ for ξ in a neighborhood of $\overline{D}'$. Then

$$(\lambda I - T)f(T) = E_\sigma, \qquad f(T)E_\sigma = f(T).$$

Hence if $x \in \mathfrak{X}_\sigma$ we have $y = f(T)x \in \mathfrak{X}_\sigma$ and $(\lambda I - T)y = x$, and if $(\lambda I - T)u = x$ with $u \in \mathfrak{X}_\sigma$ we have $y = f(T)x = f(T)(\lambda I - T)u = E_\sigma u = u$, so that $(\lambda I - T)$ is a 1-1 map of $\mathfrak{X}_\sigma$ into all of itself. Thus when T is considered as an operator in $\mathfrak{X}_\sigma$ its spectrum is contained in σ. Now if $\lambda \in \sigma$ then $\lambda \notin \sigma'$ and by what has just been proved we have $(\lambda I - T)$ a 1-1 map of $\mathfrak{X}_{\sigma'}$ into all of itself. If $(\lambda I - T)$ were also a 1-1 map of $\mathfrak{X}_\sigma$ into all of itself we see from the equation $\mathfrak{X} = \mathfrak{X}_\sigma \oplus \mathfrak{X}_{\sigma'}$ that $(\lambda I - T)$ would be a 1-1 map of $\mathfrak{X}$ into all of itself, which is impossible since $\lambda \in \sigma$. Thus when T is considered as an operator in $\mathfrak{X}_\sigma$ it has σ for its spectrum. Next let $\lambda \in \sigma$ be in the point spectrum of T. Let $x \neq 0$ be such that $(\lambda I - T)x = 0$ so that $(\lambda I - T)E_\sigma x = 0$, $(\lambda I - T)E_{\sigma'}x = 0$. Since $(\lambda I - T)$ is 1-1 on $X_{\sigma'}$ we have $E_{\sigma'}x = 0$, and hence $x = E_\sigma x$ and λ is in the point spectrum of T when T is considered as operating in $\mathfrak{X}_\sigma$. The converse is obvious. Now for any $\lambda \in \sigma$ we have

$$(\lambda I - T)\mathfrak{X} = (\lambda I - T)\mathfrak{X}_\sigma \oplus (\lambda I - T)\mathfrak{X}_{\sigma'} = (\lambda I - T)\mathfrak{X}_\sigma \oplus \mathfrak{X}_{\sigma'}$$

and so $(\lambda I - T)$ is dense in but not equal to $\mathfrak{X}$ if and only if $(\lambda I - T)\mathfrak{X}_\sigma$ is dense in but not equal to $\mathfrak{X}_\sigma$. Thus for $\lambda \in \sigma$ either of the following statements is true if and only if it is true when T is regarded as operating in $\mathfrak{X}_\sigma$: λ is in the continuous spectrum of T, λ is in the residual spectrum of T. Finally let λ be an isolated point in σ. The Laurent expansion of $R_\xi(T)$ about the point λ is

$$R_\xi(T) = \sum_{n=-\infty}^{+\infty} a_n(\lambda - \xi)^n, \qquad a_n = -\frac{1}{2\pi i}\int_C \frac{R_\xi(T)\,d\xi}{(\lambda - \xi)^{n+1}},$$

where C is a small circle about λ. Thus

$$(\lambda I - T)^k E_\lambda = \frac{1}{2\pi i}\int_C (\lambda - \xi)^k R_\xi(T)\,d\xi = -a_{-(k+1)},$$

and hence λ is a pole of order ν if and only if

$$(\lambda I - T)^\nu E_\lambda = 0, \qquad (\lambda I - T)^{\nu-1} E_\lambda \neq 0.$$

Thus λ will be a pole of order ν when T is regarded as operating in $\mathfrak{X}_\sigma$ if and only if

$$(\lambda I - T)^\nu E_\lambda E_\sigma = 0, \qquad (\lambda I - T)^{\nu-1} E_\lambda E_\sigma \neq 0,$$

but since $\lambda \in \sigma$ we have $E_\lambda E_\sigma = E_\lambda$ and hence the two conditions are equivalent. This completes the proof of the theorem.

The following corollaries will be used frequently.

2.13. THEOREM. *If σ is a spectral set of T and $\lambda \notin \sigma$ then $(\lambda I - T)$ is a* 1-1 *bicontinuous map of $\mathfrak{X}_\sigma$ into all of itself.*

2.14. THEOREM. *If σ is a spectral set of T then $E_\sigma[T] = I$ if and only if $\sigma = \sigma(T)$, and $E_\sigma[T] = 0$ if and only if σ is empty.*

This is an immediate consequence of 2.13.

2.15. THEOREM. *If λ is an isolated point in the spectrum of T then λ is a pole of order ν if and only if*

$$(\lambda I - T)^\nu E_\lambda = 0, \qquad (\lambda I - T)^{\nu-1} E_\lambda \neq 0.$$

This was established in the proof of 2.12.

2.16. DEFINITION. *As in the finite case we say that a complex number λ has index ν (a positive integer or zero) in case $(\lambda I - T)^{\nu+1} x = 0$ implies $(\lambda I - T)^\nu x = 0$ and there is an $x \in \mathfrak{X}$ with $(\lambda I - T)^\nu x = 0$, $(\lambda I - T)^{\nu-1} x \neq 0$. If no such integer ν exists we say that λ is of infinite index. For any complex number λ and integer $n \geqq 0$ we define*

$$\mathfrak{N}_\lambda^n = (\lambda I - T)^n \mathfrak{X}, \qquad \mathfrak{M}_\lambda^n = \mathop{\mathcal{E}}_{x \in \mathfrak{X}} [(\lambda I - T)^n x = 0].$$

2.17. THEOREM. *If λ is a pole of order ν for $R_\xi(T)$ then λ has index ν and $\mathfrak{X}_\lambda = \mathfrak{M}_\lambda^\nu$, $\mathfrak{X}_{\sigma'} = \mathfrak{N}_\lambda^\nu$, where σ' is the spectral set of T complementary to λ. Furthermore $\mathfrak{N}_\lambda^n = \mathfrak{N}_\lambda^\nu$ for $n \geqq \nu$ while for $1 \leqq n \leqq \nu$, $\mathfrak{N}_\lambda^n$ is a proper subset of $\mathfrak{N}_\lambda^{n-1}$.*

By 2.15 we have

$$(*) \qquad (\lambda I - T)^\nu E_\lambda = 0, \qquad (\lambda I - T)^{\nu-1} E_\lambda \neq 0.$$

Thus there is an x with $(\lambda I - T)^\nu E_\lambda x = 0$, $(\lambda I - T)^{\nu-1} E_\lambda x \neq 0$, and so the index of λ is at least ν. Now suppose that for some n and x we have $(\lambda I - T)^n x = 0$. Then since

$$R_\xi(T) = -\sum_{n=0}^{\infty} \frac{(\lambda I - T)^n}{(\lambda - \xi)^{n+1}}, \quad |\lambda - \xi| > |\lambda I - T|,$$

we see that the function

$$R_\xi(T)x = -\sum_{j=0}^{n-1} \frac{(\lambda I - T)^j}{(\lambda - \xi)^{j+1}} x$$

is regular over the entire plane except perhaps at the point $\xi=\lambda$. Thus if K is any contour surrounding $\sigma(T)$ and C is a small circle about λ we have

$$(**) \qquad x = \frac{1}{2\pi i}\int_K R_\xi(T)x\,d\xi = \frac{1}{2\pi i}\int_C R_\xi(T)x\,d\xi = E_\lambda x,$$

and hence $(\lambda I-T)^\nu x=(\lambda I-T)^\nu E_\lambda x=0$, which proves that λ has index at most ν. Thus λ has index ν. Equation (*) shows that $\mathfrak{X}_\lambda \subset \mathfrak{M}_\lambda^\nu$ and (**), which holds for $x\in\mathfrak{M}_\lambda^\nu$, shows that $\mathfrak{M}_\lambda^\nu \subset \mathfrak{X}_\lambda$. Thus $\mathfrak{M}_\lambda^\nu = \mathfrak{X}_\lambda$. Now

$$\mathfrak{X} = \mathfrak{X}_\lambda \oplus \mathfrak{X}_{\sigma'} = \mathfrak{M}_\lambda^\nu \oplus \mathfrak{X}_{\sigma'},$$

and by 2.13

$$\mathfrak{X}_{\sigma'} = (\lambda I - T)^\nu \mathfrak{X}_{\sigma'} \subset (\lambda I - T)^\nu \mathfrak{X} = \mathfrak{N}_\lambda^\nu$$

so that $\mathfrak{X}=\mathfrak{M}_\lambda^\nu+\mathfrak{N}_\lambda^\nu$. Since λ has index ν we have $\mathfrak{M}_\lambda^\nu\cdot\mathfrak{N}_\lambda^\nu=0$ and hence $\mathfrak{X}=\mathfrak{M}_\lambda^\nu\oplus\mathfrak{N}_\lambda^\nu$. Thus $\mathfrak{X}_{\sigma'}$ and $\mathfrak{N}_\lambda^\nu$ are both complements to $\mathfrak{M}_\lambda^\nu$ and $\mathfrak{X}_{\sigma'}\subset\mathfrak{N}_\lambda^\nu$ which proves that $\mathfrak{N}_\lambda^\nu=\mathfrak{X}_{\sigma'}$. By 2.13 $\mathfrak{N}_\lambda^{\nu+k}=(\lambda I-T)^k\mathfrak{N}_\lambda^\nu=(\lambda I-T)^k\mathfrak{X}_{\sigma'}=\mathfrak{X}_{\sigma'}=\mathfrak{N}_\lambda^\nu$ and so $\mathfrak{N}_\lambda^n=\mathfrak{N}_\lambda^\nu$ for $n\geqq\nu$. Since if, for some k, $\mathfrak{N}_\lambda^{k+1}=\mathfrak{N}_\lambda^k$ we have $\mathfrak{N}_\lambda^n=\mathfrak{N}_\lambda^k$ for all $n\geqq k$, to prove the final assertion of the theorem it will suffice to show that $\mathfrak{N}_\lambda^\nu$ is a proper subset of $\mathfrak{N}_\lambda^{\nu-1}$. Since λ has index ν there is an x with $(\lambda I-T)^\nu x=0$, $0\neq y=(\lambda I-T)^{\nu-1}x\in\mathfrak{N}_\lambda^{\nu-1}$. Since $0\neq y\in\mathfrak{M}_\lambda^\nu$ and $\mathfrak{M}_\lambda^\nu\mathfrak{N}_\lambda^\nu=0$ we have $y\in\mathfrak{N}_\lambda^{\nu-1}-\mathfrak{N}_\lambda^\nu$ which completes the proof of the theorem.

There is a converse to Theorem 2.17. It states that if for some λ and m we have $\mathfrak{N}_\lambda^m\oplus\mathfrak{M}_\lambda^m=\mathfrak{X}$ and $\mathfrak{N}_\lambda^m$ closed, then λ is either in the resolvent set $\rho(T)$ or else a pole of order at most m for $R_\xi(T)$. This result is a special case of 2.23 which will be proved later.

2.18. THEOREM. *Let λ be either in $\rho(T)$ or an isolated point of the spectrum. Let δ be the distance from λ to the rest of the spectrum. Then the following statements are equivalent*:

(1) $x\in\mathfrak{X}_\lambda$.

(2) $\lim_n (\lambda I-T)^n\epsilon^{-n}x=0$ *for every* $\epsilon>0$.

(3) *For some positive* $\epsilon<\delta$ *the equation* (2) *is satisfied.*

To see that (1) implies (2) let C be the circle $|\lambda-\xi|=\rho$ where $0<\rho<\epsilon$. Then

$$(\lambda I - T)^n\epsilon^{-n}E_\lambda = \frac{1}{2\pi i}\int_C (\lambda - \xi)^n\epsilon^{-n}R_\xi(T)d\xi \to 0.$$

Conversely suppose (2) is satisfied for some positive $\epsilon < \delta$. Let $0 < \epsilon_1 < \epsilon < \epsilon_2 < \epsilon_3$, and C_i $(i = 1, 2, 3)$, the circle $|\lambda - \xi| = \epsilon_i$, be chosen so that the spectrum $\sigma(T)$ with the exception of the point $\xi = \lambda$ is in the open annulus bounded by C_2 and C_3. Then the function $f_n(\xi) = [1 - (\lambda - \xi)^n\epsilon^{-n}]^{-1}$ is in $\mathcal{J}(T)$ and

$$f_n(T) = \frac{1}{2\pi i}\left[\int_{C_1} f_n(\xi)R_\xi(T)d\xi - \int_{C_2} f_n(\xi)R_\xi(T)d\xi + \int_{C_3} f_n(\xi)R_\xi(T)d\xi\right],$$

$$f_n(T) \to E_\lambda.$$

Since we are assuming (2) to hold for ϵ we have $f_n^{-1}(T)x = x - (\lambda I - T)^n\epsilon^{-n}x \to x$ and thus for arbitrary $\rho > 0$, $|x - f_n^{-1}(T)x| < \rho$, $|f_n(T)x - x| < \rho|f_n(T)|$ for all large n. Since $f_n(T) \to E_\lambda$ we have $|E_\lambda x - x| < \rho|E_\lambda|$ and since ρ is arbitrary, $E_\lambda x = x$.

2.19. THEOREM (THE MINIMAL EQUATION THEOREM). *If $f \in \mathcal{J}(T)$ then $f(T) = 0$ if and only if*

(1) *For every pole λ of $R_\xi(T)$ of order ν*

$$f^{(j)}(\lambda) = 0, \qquad j = 0, 1, \cdots, \nu - 1,$$

(2) *$f(\lambda) \equiv 0$ in a neighborhood of the spectrum $\sigma(T)$ excluding poles of $R_\xi(T)$.*

Suppose (1) and (2) are satisfied. Let $D = D_1 + \cdots + D_k$ be a domain upon which f is regular and single-valued and with $D \supset \sigma(T)$. Clearly, in view of (2),

$$(*) \qquad f(T) = \frac{1}{2\pi i}\int_{C^*} f(\lambda)R_\lambda(T)d\lambda,$$

where C^* is the boundary of those D_i containing no singularities of $R_\xi(T)$ except a finite number of poles. Let $\lambda_1, \cdots, \lambda_k$ be the poles contained within C^* and let $\nu_1, \cdots, \nu_k$ be their orders. Then from (1) we have $f(\lambda) = (\lambda_i - \lambda)^{\nu_i}g_i(\lambda)$ where g_i is regular at λ_i and $g_i \in \mathcal{J}(T)$. Hence by 2.15, $f(T)E_{\lambda_i} = g_i(T)(\lambda_i I - T)^{\nu_i}E_{\lambda_i} = 0$ and hence from (*) we have

$$f(T) = \sum_{i=1}^{k} f(T)E_{\lambda_i} = 0.$$

Now conversely suppose $f(T) = 0$. Then $0 = \sigma(f(T)) = f(\sigma(T))$ and so $f(\lambda)$ is identically zero in any connected domain (upon which it is regular) containing infinitely many points of $\sigma(T)$. Hence equation (*) holds where C^* is now the boundary of a finite number of domains containing only a finite number of singularities of $R_\xi(T)$. Let $\lambda_1, \cdots, \lambda_n$ be the poles within C^* and

$\nu_1, \cdots, \nu_n$ their orders. Let $\lambda_{n+1}, \cdots, \lambda_m$ be the essential singularities within C^*. Now suppose that for some $i \leqq n$ we have

$$f^{(k)}(\lambda_i) \neq 0, \qquad f^{(j)}(\lambda_i) = 0, \qquad j = 0, \cdots, k-1,$$

where $k < \nu_i$. Define a function $g \in \mathcal{J}(T)$ so that in a neighborhood of λ_i we have $f(\lambda) = (\lambda_i - \lambda)^k g(\lambda)$ while $g(\lambda) \equiv 1$ in a neighborhood of the spectrum $\sigma(T)$ excluding λ_i. Thus g, g^{-1} are both in $\mathcal{J}(T)$. Since $k < \nu_i$ we have by 2.17 a point $x \in \mathfrak{X}_{\lambda_i}$ with $(\lambda_i I - T)^k x \neq 0$, and thus since $g(T)$ has an inverse

$$f(T)x = f(T)E_{\lambda_i}x = g(T)(\lambda_i I - T)^k E_{\lambda_i}x \neq 0,$$

which is a contradiction.

The coefficient a_{-n-1} of $(\lambda_i - \xi)^{-n-1}$ in the Laurent expansion of $R_\xi(T)$ about the point $\xi = \lambda_i$ is (as was shown in the proof of 2.12) $-(\lambda_i I - T)^n E_{\lambda_i}$ and thus for $i > n$ we have $(\lambda_i I - T)^n E_{\lambda_i} \neq 0$ for all $n = 1, 2, \cdots$. Just as before suppose that for some $i > n$ and some $k \geqq 1$ we have $f^{(k)}(\lambda_i) \neq 0$, $f^{(j)}(\lambda_i) = 0, j = 0, \cdots, k-1$, and as before define $g \in \mathcal{J}(T)$ so that $g^{-1}(T)$ exists and $f(T)E_{\lambda_i} = (\lambda_i I - T)^k g(T) E_{\lambda_i} \neq 0$, which again is a contradiction.

An immediate corollary is

2.20. THEOREM. *If $f, g \in \mathcal{J}(T)$ then $f(T) = g(T)$ if and only if*

(1) *For every pole λ of $R_\xi(T)$ of order ν*

$$f^{(j)}(\lambda) = g^{(j)}(\lambda), \qquad j = 0, \cdots, \nu - 1.$$

(2) *$f(\lambda) = g(\lambda)$ for every λ in a neighborhood of the spectrum $\sigma(T)$ excluding poles.*

2.21. THEOREM. *Let $\lambda_1, \cdots, \lambda_k$ be poles of $R_\xi(T)$ of orders $\nu_1, \cdots, \nu_k$ respectively. Let σ' be the complement of the spectral set $\sigma = (\lambda_1, \cdots, \lambda_k)$. Then for every $f \in \mathcal{J}(T)$ we have*

$$f(T) = f(T)E_{\sigma'} + \sum_{i=1}^{k} \sum_{n=0}^{\nu_i - 1} \frac{(T - \lambda_i I)^n}{n!} f^{(n)}(\lambda_i) E_{\lambda_i}.$$

This is an immediate corollary of 2.20.

2.22. DEFINITION. *Let $f \in \mathcal{J}(T)$, then as in 1.1 we define the manifolds*

$$\mathfrak{N}[f] = f(T)\mathfrak{X}, \qquad \mathfrak{M}[f] = \mathop{\mathcal{E}}_{x \in \mathfrak{X}} [f(T)x = 0].$$

2.23. THEOREM. *Let $f \in \mathcal{J}(T)$ and suppose that $f(\lambda)$ is not identically zero on any of the domains in which it is regular. Let $\lambda_1, \cdots, \lambda_k$ be the roots of $f(\lambda)$ and $m_1, \cdots, m_k$ their multiplicities. Then the following statements are equivalent:*

(1) *$\mathfrak{M}[f] \oplus \mathfrak{N}[f] = \mathfrak{X}$, $\mathfrak{N}[f]$ is closed.*

(2) *For* $i=1, \cdots, k$, λ_i *is either in the resolvent set* $\rho(T)$ *or else a pole of order* $\nu_i \leqq m_i$ *of* $R_\xi(T)$.

(3) *The finite set* $\sigma=(\lambda_1, \cdots, \lambda_k)\sigma(T)$ *is a spectral set of* T *and*

$$\mathfrak{X}_\sigma = \mathfrak{M}[f], \qquad \mathfrak{X}_{\sigma'} = \mathfrak{N}[f].$$

Assuming (1) we have continuous projections E, E' with $E\mathfrak{X}=\mathfrak{M}[f]$, $E'\mathfrak{X}=\mathfrak{N}[f]$, $EE'=0$, $I=E+E'$, $TE=ET$. Also for any λ we have $(\lambda I-T)$ a 1-1 map of $\mathfrak{X}$ into all of itself if and only if $(\lambda I-T)$ is a 1-1 map of $E\mathfrak{X}$ into all of itself and of $E'\mathfrak{X}$ into all of itself. Thus the spectrum $\sigma(T)$ of T is $\sigma_1+\sigma_1'$ where σ_1, σ_1' are the spectrums of T when it is considered as an operator in $E\mathfrak{X}$, $E'\mathfrak{X}$ respectively. In $E\mathfrak{X}$ we have $f(T)=0$ and so σ_1 is contained in the point set $(\lambda_1, \cdots, \lambda_k)$. Since $f(T)E=0$ we have $f(T)=f(T)E'$ and hence

$$\mathfrak{N}[f] = f(T)\mathfrak{X} = f(T)E'\mathfrak{X} = f(T)\mathfrak{N}[f] \subset f(T)\mathfrak{X} = \mathfrak{N}[f].$$

Therefore, since $\mathfrak{M}[f]\cdot\mathfrak{N}[f]=0$, $f(T)$ is a 1-1 map of $\mathfrak{N}[f]=E'\mathfrak{X}$ into all of itself. Hence when regarded as an operator in $E'\mathfrak{X}$, $0\in\rho(f(T))$ and hence $\lambda_i\in\rho(T)$, $i=1, \cdots, k$. Thus σ_1' is a closed point set not containing $\sigma=\sigma(T)(\lambda_1, \cdots, \lambda_k)$. This shows that σ as well as $\lambda_i\in\sigma$ is a spectral set of T. Now let $x=Ex+E'x\in X_{\lambda_i}$, then 2.18 gives

$$\lim_n (\lambda_i I - T)^n \epsilon^{-n} E'x = 0, \qquad \epsilon > 0,$$

and hence since $(\lambda_i I-T)$ is a 1-1 bicontinuous map of $E'\mathfrak{X}$ into all of itself we have $E'x=0$ and $x=Ex\in\mathfrak{M}[f]$. Thus

$$\mathfrak{X}_\sigma = \mathfrak{X}_{\lambda_1} \oplus \cdots \oplus \mathfrak{X}_{\lambda_k} \subset \mathfrak{M}[f].$$

On the other hand if $x\in\mathfrak{M}_{\lambda_i}^{m_i}$ we have (by 2.18) $x\in\mathfrak{X}_{\lambda_i}$ and so by 1.4

$$\mathfrak{M}[P] = \mathfrak{M}_{\lambda_1}^{m_1} \oplus \cdots \oplus \mathfrak{M}_{\lambda_k}^{m_k} \subset \mathfrak{X}_\sigma,$$

where $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{m_i}$. If $h(\lambda)$ is defined on the same domain as $f(\lambda)$ by the equation $f(\lambda)=P(\lambda)\cdot h(\lambda)$ we see that $h(\lambda)$ has no roots in $\sigma(T)$ and hence $h^{-1}(T)$ exists, which shows that $\mathfrak{N}[P]=\mathfrak{M}[f]$ and thus the above equations give $\mathfrak{M}[f]=\mathfrak{X}_\sigma$. Now since $f(\lambda)\not\equiv 0$ and in $\mathfrak{X}_\sigma=\mathfrak{M}[f]$ we have $f(T)=0$ we see from the minimal equation theorem that in $\mathfrak{X}_\sigma$ every spectral point λ_i must be a pole of order $\nu_i\leqq m_i$. Thus (2) follows from 2.12. It has therefore been proved that (1) implies (2). During the course of this proof all but the last part of (3) has also been established. To see that $\mathfrak{X}_{\sigma'}=\mathfrak{N}[f]$ note first that since $0\notin f(\sigma')$ we have (by 2.12)

$$\mathfrak{X}_{\sigma'} = f(T)\mathfrak{X}_{\sigma'} \subset f(T)\mathfrak{X} = \mathfrak{N}[f].$$

Next

$$\mathfrak{X} = \mathfrak{X}_\sigma \oplus \mathfrak{X}_{\sigma'} = \mathfrak{M}[f] \oplus \mathfrak{X}_{\sigma'},$$

and so $\mathfrak{N}[f]$ and $\mathfrak{X}_{\sigma'}$ are both complements of $\mathfrak{M}[f]$ and $\mathfrak{X}_{\sigma'} \subset \mathfrak{N}[f]$. This clearly implies that $\mathfrak{X}_{\sigma'} = \mathfrak{N}[f]$. Thus (1) implies (3). Clearly (3) implies (1). Now assume (2) and let $\nu_i = 0$ if λ_i is in $\rho(T)$, then if $P(\lambda)$ is the polynomial defined above we have by 2.17 and 1.4

$$\begin{aligned}\mathfrak{X}_\sigma &= \mathfrak{X}_{\lambda_1} \oplus \cdots \oplus \mathfrak{X}_{\lambda_k} = \mathfrak{M}_{\lambda_1}^{\nu_1} \oplus \cdots \oplus \mathfrak{M}_{\lambda_k}^{\nu_k} \\ &= \mathfrak{M}_{\lambda_1}^{m_1} \oplus \cdots \oplus \mathfrak{M}_{\lambda_k}^{m_k} = \mathfrak{M}[P] = \mathfrak{M}[f].\end{aligned}$$

Also by 2.17 and 1.4 we have, for every x such that $P^2(T)x = 0$,

$$x \in \mathfrak{M}_{\lambda_1}^{2m_1} \oplus \cdots \oplus \mathfrak{M}_{\lambda_k}^{2m_k} = \mathfrak{M}_{\lambda_1}^{\nu_1} \oplus \cdots \oplus \mathfrak{M}_{\lambda_k}^{\nu_k} = \mathfrak{M}[P]$$

and so $P(T)x = 0$, that is $\mathfrak{N}[P] \cdot \mathfrak{M}[P] = 0$. Thus since (using 2.12)

$$\mathfrak{N}[f] = \mathfrak{N}[P] \supset P(T)\mathfrak{X}_{\sigma'} = \mathfrak{X}_{\sigma'}$$

we have

$$\mathfrak{X} = \mathfrak{X}_\sigma \oplus \mathfrak{X}_{\sigma'} = \mathfrak{M}[P] \oplus \mathfrak{N}[P] = \mathfrak{M}[f] \oplus \mathfrak{N}[f],$$

and $\mathfrak{X}_{\sigma'} = \mathfrak{N}[f]$ is closed. This completes the proof of 2.23.

The formula $\sigma(f(T)) = f(\sigma(T))$ may be given a much sharper form. To see this let us first examine the finite case of an $n \times n$ matrix T whose distinct characteristic numbers are $\lambda_1, \cdots, \lambda_k$. The formula asserts that $f(\lambda_1), \cdots, f(\lambda_k)$ are the characteristic numbers of $f(T)$. However a theorem due to Sylvester asserts more; namely, if $\lambda_1, \cdots, \lambda_n$ are the characteristic numbers of T, each repeated according to its multiplicity, then $f(\lambda_1), \cdots, f(\lambda_n)$ are the characteristic numbers of $f(T)$, where the number of repetitions of a given number in this array is its multiplicity. In the language of determinants this theorem of Sylvester asserts that if $\lambda_1, \cdots, \lambda_k$ are the *distinct* characteristic numbers of T and if m_i is the multiplicity of λ_i then $\mathrm{Det}(f(T) - \lambda I) = \prod_{i=1}^{k} (f(\lambda_i) - \lambda)^{m_i}$. Recalling that the multiplicity $m(\lambda_i, T)$ of a root λ_i of T of index ν_i is the number of linearly independent solutions of $(\lambda_i I - T)^{\nu_i} x = 0$, that is $m(\lambda_i, T)$ is the dimension of $\mathfrak{M}_{\lambda_i}^{m_i} = \mathfrak{M}_{\lambda_i}^{\nu_i}$, it is seen that the Sylvester theorem, that is the formula

$$\text{(i)} \qquad m(\mu, f(T)) = \sum_{\lambda \in f^{-1}(\mu)} m(\lambda, T)$$

(where $m(\mu, f(T))$ is the multiplicity of μ as a root of $f(T)$ and the sum on the right is taken over all λ such that $f(\lambda) = \mu$) is an immediate consequence of

$$\text{(ii)} \qquad \mathfrak{X}_\mu[f(T)] = \sum_{\lambda \in f^{-1}(\mu)} \mathfrak{X}_\lambda[T]$$

(where the sum is a vector sum). Formula (ii) which states that two manifolds are equal is for many purposes more desirable than (i) which merely states that the dimensions of these manifolds are the same. On the other

hand (ii) is readily derived from (i) as follows. Suppose $(\lambda I-T)^r x=0$ (that is $x\in\mathfrak{X}_\lambda(T)$), where $f(\lambda)=\mu$, then $(\mu I-f(T))^r x=(f(\lambda)I-f(T))^r x=(\lambda I-T)^r g^r(T)x=0$ and so $x\in\mathfrak{X}_\mu[f(T)]$. Thus the left side of (ii) contains the right but by (i) the dimension of the left side is the same as the dimension of the right side and so the manifolds are the same. A form of this theorem which is more desirable than (ii) is

$$\text{(iii)}\qquad E_\mu[f(T)]=\sum_{\lambda\in f^{-1}(\mu)}E_\lambda[T],$$

which states the equality of two projections rather than (as (ii) does) the equality of the manifolds upon which they project.

In Theorems 2.24 and 2.29 to follow we shall give two forms of (iii) applicable to a general linear operator T.

2.24. THEOREM. *Let $f\in\mathcal{J}(T)$ and let τ be a spectral set of $f(T)$. Then $\sigma=\sigma(T)\cdot f^{-1}(\tau)$ is a spectral set of T and*

$$E_\tau[f(T)]=E_\sigma[T].$$

Let τ' be the complement of τ in $\sigma(f(T))$. Then since $\sigma(f(T))=f(\sigma(T))$ we have $\sigma+\sigma'=\sigma(T)$ where $\sigma'=\sigma(T)\cdot f^{-1}(\tau)$. Since τ, τ' are closed and disjoint and $f(\lambda)$ is continuous we see that σ, σ' are closed and disjoint and thus they are complementary spectral sets of T. Let D, D' be $f(T)$-admissible domains with $\overline{D}\overline{D}'=0$, $\tau=\sigma(f(T))\cdot D$, $\tau'=\sigma(f(T))D'$. Then there are T-admissible domains D_1, D_1' with $\overline{D}_1\overline{D}_1'=0$, $\sigma=\sigma(T)D_1$, $\sigma'=\sigma(T)D_1'$, f is regular on $\overline{D}_1+\overline{D}_1'$ and $f(\overline{D}_1)\subset D$, $f(\overline{D}_1')\subset D'$. Let g, h be two characteristic functions with $g\in\mathcal{J}(f(T))$, $h\in\mathcal{J}(T)$ and such that

$$g(\xi)=\begin{cases}1, & \xi\in\overline{D},\\ 0, & \xi\in\overline{D}',\end{cases}\qquad h(\xi)=\begin{cases}1, & \xi\in\overline{D}_1,\\ 0, & \xi\in\overline{D}_1',\end{cases}$$

and let $F(\xi)=g(f(\xi))$. By 2.10 we have $F(T)=g(f(T))$. Since $F(\xi)=h(\xi)$ for $\xi\in\overline{D}_1+\overline{D}_1'$ we have $F(T)=h(T)$ and

$$E_\tau[f(T)]=g(f(T))=F(T)=h(T)=E_\sigma[T].$$

2.25. LEMMA. *Let Γ be a set of complex numbers, and for each $\xi\in\Gamma$ let $B(\xi)$, $A_n(\xi)$, $n=1, 2, \cdots$, be bounded linear operators which have inverses. Let M, M_1, M_2, $\cdots$ be constants such that*

$$|B^{-1}(\xi)|\leqq M,\qquad |A_n^{-1}(\xi)|\leqq M_n,\qquad \xi\in\Gamma.$$

Then if $\lim_n A_n(\xi)=B(\xi)$ uniformly for $\xi\in\Gamma$ we have also $\lim_n A_n^{-1}(\xi)=B^{-1}(\xi)$ uniformly for $\xi\in\Gamma$.

The proof is the same as that given by Gelfand(9) for a special case.

(9) Loc. cit.

We shall prove the lemma first for the case $B(\xi)\equiv I$. In this case there is an integer n_0 such that $|A_n(\xi)-I|<1/2$ for $\xi\in\Gamma$ and $n\geqq n_0$. Thus for $n\geqq n_0$

$$A_n^{-1}(\xi)=\sum_{p=0}^{\infty}[I-A_n(\xi)]^p, \qquad |A_n^{-1}(\xi)|\leqq 2, \qquad \xi\in\Gamma.$$

Hence $\text{l.u.b.}_{1\leqq n<\infty}\ \text{l.u.b.}_{\xi\in\Gamma}|A_n^{-1}(\xi)|=K<\infty$ and

$$|A_n^{-1}(\xi)-I|=|A_n^{-1}(\xi)(I-A_n(\xi))|\leqq K|I-A_n(\xi)|,$$

which proves the special case of $B(\xi)\equiv I$. Now in the general case we have

$$|A_n(\xi)B^{-1}(\xi)-I|=|(A_n(\xi)-B(\xi))B^{-1}(\xi)|\leqq M|A_n(\xi)-B(\xi)|,$$

and so by the first case we have

$$\lim_n B(\xi)A_n^{-1}(\xi)=I$$

uniformly on Γ. Thus

$$|A_n^{-1}(\xi)-B^{-1}(\xi)|=|B^{-1}(\xi)(I-B(\xi)A_n^{-1}(\xi))|\leqq M|I-B(\xi)A_n^{-1}(\xi)|,$$

which proves the lemma.

2.26. LEMMA. *Let B, A_n, $n=1, 2, \cdots$, be bounded linear operators and let Γ be a closed set of complex numbers with $\Gamma\subset\rho(B)\rho(A_n)$, $n=1, 2, \cdots$. Then if $A_n\to B$ we have $R_\xi(A_n)\to R_\xi(B)$ uniformly on Γ.*

Let $B(\xi)=\xi I-B$, $A_n(\xi)=\xi I-A_n$. Since Γ is closed and in $\rho(B)$ it must be at positive distance from $\sigma(B)$. Since $|R_\xi(B)|$ is continuous on Γ and $|R_\xi(B)|\to 0$ as $\xi\to\infty$ we see that $\text{l.u.b.}_{\xi\in\Gamma}|B^{-1}(\xi)|<\infty$, similarly for $A_n^{-1}(\xi)$. The lemma thus follows from the preceding one.

2.27. DEFINITION. *By $\mathfrak{R}(T)$ we shall mean the ring of all operators of the form $f(T)$ where $f\in\mathcal{F}(T)$, and by $\overline{\mathfrak{R}}(T)$ the ring of all operators which are limits (in the uniform topology of operators) of elements in $\mathfrak{R}(T)$.*

2.28. LEMMA. *With every $U\in\overline{\mathfrak{R}}(T)$ is associated a uniquely defined scalar function f called the spectral function of U, such that:*

(i) *The domain of definition of f is $\sigma(T)$ and $f(\lambda)$ is continuous for $\lambda\in\sigma(T)$.*

(ii) *If $f_n\in\mathcal{F}(T)$ and $f_n(T)\to U$ then $f_n(\lambda)\to f(\lambda)$ uniformly for $\lambda\in\sigma(T)$.*

(iii) *$f(\sigma(T))\subset\sigma(U)$.*

If $f_n\in\mathcal{F}(T)$ and $f_n(T)\to U$ then for every $\lambda\in\sigma(T)$ we have $f_n(\lambda)-f_m(\lambda)\in\sigma(f_n(T)-f_m(T))$ and so

$$\text{l.u.b.}_{\lambda\in\sigma(T)}|f_n(\lambda)-f_m(\lambda)|\leqq|f_n(T)-f_m(T)|\to 0.$$

Thus $f(\lambda)=\lim_n f_n(\lambda)$ exists on $\sigma(T)$ and is continuous there. If also $g_n(T)\to U$

then by the same argument $|f_n(\lambda)-g_n(\lambda)|\to 0$ uniformly for $\lambda\in\sigma(T)$, and so $f(\lambda)$ is independent of the particular sequence $f_n(T)\in\mathfrak{R}(T)$ used to approach U. To prove (iii) let $\lambda\in\sigma(T)$, so that $f_n(\lambda)\in\sigma(f_n(T))$, and suppose that $f(\lambda)\notin\sigma(U)$. Since $f_n(\lambda)I-f_n(T)\to f(\lambda)I-U$ and $[f(\lambda)I-U]^{-1}$ exists we have, for all large n, $[f_n(\lambda)-f_n(T)]^{-1}$ existing, which contradicts the fact that $f_n(\lambda)\in\sigma(f_n(T))$.

We are now prepared to give the general form of the Sylvester theorem.

2.29. THEOREM. *Let f be the spectral function of an operator $U\in\overline{\mathfrak{R}}(T)$ and let τ be a spectral set of U. Then $\sigma=f^{-1}(\tau)$ is a spectral set of T and*

$$E_\sigma[T] = E_\tau[U].$$

Let τ' be the spectral set of U complementary to τ and let $\sigma'=f^{-1}(\tau')$. By the preceding lemma $f(\sigma(T))\subset\tau+\tau'$ and so $\sigma+\sigma'=\sigma(T)$, $\sigma\sigma'=0$. Since τ, τ' are both closed in $\sigma(U)$ and f is continuous we see that σ, σ' are both closed and hence are complementary spectral sets of T. Let D, D' be U-admissible domains with $\tau=D\cdot\sigma(U)$, $\tau'=D'\cdot\sigma(U)$, $\overline{DD}'=0$ and let C be the boundary of D. Let $f_n\in\mathcal{J}(T)$ and $f_n(T)\to U$. Since $f_n(\lambda)\to f(\lambda)$ uniformly on $\sigma(T)$ and $f(\sigma(T))\subset\tau+\tau'$ we have for all large n

$$(*) \qquad f_n(\sigma)\subset D, \qquad f_n(\sigma')\subset D'.$$

Thus $f_n(\sigma)=f_n(\sigma)D$ is a spectral set of $f_n(T)$ and $\sigma(T)f_n^{-1}(f_n(\sigma))=\sigma$, and so, by 2.24, $E_\sigma[T]=E_{f_n(\sigma)}[f_n(T)]$. Now (*) shows that $C\subset\rho[f_n(T)]$ and thus, by 2.26, $R_\xi(f_n(T))\to R_\xi(U)$ uniformly for $\xi\in C$. Hence

$$E_\sigma[T] = E_{f_n(\sigma)}[f_n(T)] = \frac{1}{2\pi i}\int_C R_\xi(f_n(T))\,d\xi \to \frac{1}{2\pi i}\int_C R_\xi(U)\,d\xi = E_\tau[U],$$

which completes the proof of the theorem.

2.30. THEOREM. *Let $\mu\in\sigma(f(T))$ where $f\in\mathcal{J}(T)$ and suppose that $f(\lambda)\not\equiv\mu$ on any of the domains in which it is regular. Let $\lambda_1,\cdots,\lambda_k$ be those points λ in $\sigma(T)$ where $f(\lambda)=\mu$. Then*

(1) *If μ is a pole of $R_\xi(f(T))$ then $\lambda_1,\cdots,\lambda_k$ are poles of $R_\xi(T)$.*

(2) *Conversely if $\lambda_1,\cdots,\lambda_k$ are poles of $R_\xi(T)$ then μ is a pole of $R_\xi(f(T))$.*

Since $f(\lambda)$ is not identically μ on any of the domains in which it is regular it is clear that there are only a finite number $\lambda_1,\cdots,\lambda_k$ of roots of $f(\lambda)-\mu=0$ which lie in the spectrum $\sigma(T)$. Since $f(\sigma(T))=\sigma(f(T))$ and $\mu\in\sigma(f(T))$ there is at least one root of $f(\lambda)-\mu=0$ in $\sigma(T)$. Let $m_1,\cdots,m_k$ be the multiplicities of $\lambda_1,\cdots,\lambda_k$ as roots of $f(\lambda)-\mu=0$ and let $h\in\mathcal{J}(T)$ be defined on the same domain as f by the equation

$$\mu - f(\lambda) = (\lambda_1-\lambda)^{m_1}\cdots(\lambda_k-\lambda)^{m_k}h(\lambda)$$

so that $h(\lambda)$ has no roots in $\sigma(T)$ and thus $h^{-1}(T)$ exists. To prove (1) we have,

by 2.24, $\sigma=(\lambda_1, \cdots, \lambda_k)$ a spectral set of T and

$$(*) \qquad E_\mu[f(T)] = E_{\lambda_1}[T] + \cdots + E_{\lambda_k}[T].$$

Now for any integer n we have

$$(**) \qquad \begin{aligned}[\mu I - f(T)]^n E_\mu[f(T)] \\ = (\lambda_1 I - T)^{m_1 n} \cdots (\lambda_k I - T)^{m_k n} h^n(T)(E_{\lambda_1}[T] + \cdots + E_{\lambda_k}[T]).\end{aligned}$$

If $n-1$ is the order of μ as a pole of $R_\xi(f(T))$ we have, by 2.15, the left side of (**) vanishing, and upon operating with $E_{\lambda_i}[T]$ we get

$$(\lambda_1 I - T)^{m_1 n} \cdots (\lambda_k I - T)^{m_k n} h^n(T) E_{\lambda_i}[T] = 0.$$

Since $h^{-1}(T)$ exists and since, by 2.13, $(\lambda_j I - T)$ for $j \neq i$ is a 1-1 map of $\mathfrak{X}_{\lambda_i}$ into all of itself we have

$$(\lambda_i I - T)^{m_i n} E_{\lambda_i}[T] = 0,$$

which, since λ_i is isolated in $\sigma(T)$, proves (by 2.15) that λ_i is a pole of $R_\xi(T)$. Now conversely suppose that λ_i, $i=1, \cdots, k$, is a pole of order ν_i for $R_\xi(T)$ and let n be an integer with $n \geqq \nu_i$, $i=1, \cdots, k$. Since $f(\lambda)$ is continuous on $\sigma(T)$ it is seen that μ is isolated in $\sigma[f(T)]$ and thus (*) and hence (**) are valid. Since, by 2.15, $(\lambda_i I - T)^{m_i n} E_{\lambda_i}[T] = 0$ this same theorem shows in view of (**) that μ is a pole of $R_\xi(f(T))$.

2.31. DEFINITION. *Let S be the unit sphere in $\mathfrak{X}$. An operator U in $\mathfrak{X}$ is said to be compact in case US is compact in $\mathfrak{X}$ and it is said to be weakly compact in case US is weakly (sequentially) compact in $\mathfrak{X}$.*

We prefer the terminology of 2.31 to the usual terms, that is, completely continuous and weakly completely continuous.

2.32. THEOREM. *Let $f \in \mathcal{F}(T)$ be such that $f(T)$ is compact. Then every $\lambda \in \sigma(T)$ with $f(\lambda) \neq 0$ is a pole of $R_\xi(T)$ and the corresponding manifold $\mathfrak{X}_\lambda[T]$ has finite dimension.*

By a well known result of F. Riesz(9) $\mu = f(\lambda)$ is isolated in the spectrum $\sigma(U)$ of the compact operator $U=f(T)$. Thus if $\sigma = \sigma(T) f^{-1}(\mu)$ we have, by 2.24, σ a spectral set of T and

$$\mathfrak{X}_\mu[U] = \mathfrak{X}_\sigma[T].$$

Now by 2.13 we see that U is a 1-1 map of $\mathfrak{X}_\mu[U]$ into all of itself and hence the unit sphere in $\mathfrak{X}_\mu[U]$ is compact and $\mathfrak{X}_\mu[U]$, therefore, has finite dimension. Thus the spectrum of T where it is considered as operating in $\mathfrak{X}_\sigma[T]$ is finite and by 2.12, therefore, σ is finite, which shows that λ is a spectral set of T. Hence $\mathfrak{X}_\lambda[T] \subset \mathfrak{X}_\sigma[T]$ has finite dimension and, in $\mathfrak{X}_\lambda[T]$, $(\lambda I - T)$ has the

(9) F. Riesz, *Über lineare Funktionalgleichungen*, Acta Math. vol. 41 (1918) pp. 71–98.

single point 0 for its spectrum. Thus in $\mathfrak{X}_\lambda[T]$ the matrix $(\lambda I - T)$ is nilpotent, that is, for some n, $(\lambda I - T)^n E_\lambda[T] = 0$ which proves that λ is a pole of $R_\xi[T]$.

2.33. THEOREM. *Let $f_n \in \mathcal{J}(T)$ and let $U = \lim_n f_n(T)$ be a compact operator. Then every $\lambda \in \sigma(T)$ for which $f_n(\lambda)$ does not approach 0 is a pole of $R_\xi(T)$ and the corresponding manifold $\mathfrak{X}_\lambda[T]$ has finite dimension.*

Let $f(\lambda)$ be the spectral function of U so that $f(\lambda) = \lim_n f_n(\lambda)$ for every $\lambda \in \sigma(T)$. Thus if $\mu = f(\lambda) \neq 0$ we have, by 2.28, $\mu \in \sigma(U)$ and by Riesz's result μ is isolated in $\sigma(U)$. The proof now proceeds as it did in the preceding theorem except that we use 2.29 instead of 2.24.

2.34. THEOREM. *If T^n is compact then every spectral point $\mu \neq 0$ of T is a pole of $R_\xi(T)$.*

This is an immediate corollary of 2.32.

2.35. *The Fredholm theory.* If $U = f(T)$ is compact, where $f(\lambda)$ is regular in the circle $|\lambda| \leqq |T|$ and has no roots other than $\lambda = 0$ in this circle, then 2.32 shows that all points $\mu \neq 0$ in $\sigma(T)$ are isolated in $\sigma(T)$ and are poles of $R_\xi(T)$. In view of 2.17 it is seen that all of the principal results of the Fredholm theory are valid for the operator T. In particular if T is a weakly compact operator in a Lebesgue space L it is known (Dunford and Pettis[10] and Phillips[11]) that T^2 is compact and so the Fredholm theory is valid for weakly compact operators in L.

3. Convergence to projections.

3.1. THEOREM. *Every projection $E \in \overline{\mathfrak{R}}(T)$ is in $\mathfrak{R}(T)$ and $E = E_\sigma[T]$, where σ is a spectral set of T and consists of all $\lambda \in \sigma(T)$ where the spectral function $f(\lambda)$ of E has the value 1. Thus $E = 0$ if and only if there are no such λ.*

The proof will require the following lemma.

3.2. LEMMA. *Every $\lambda \neq 0, 1$ is in the resolvent set $\rho(E)$ of a continuous projection E and*

$$R_\xi(E) = E/(\xi - 1) + (I - E)/\xi, \qquad \xi \neq 0, 1.$$

Thus $E_1[E] = E$, $E_0[E] = I - E$, $\sigma(E) = 0$ if and only if $E = 0$; $\sigma(E) = 1$ if and only if $E = I$, and $\sigma(E) = \sigma_p(E)$.

The formula for $R_\xi(E)$ is verified by multiplying by $\xi I - E$ and the remaining conclusions follow immediately.

---

(10) N. Dunford and B. J. Pettis, *Linear operations on summable functions*, Trans. Amer. Math. Soc. vol. 47 (1940) pp. 323–392.

(11) R. S. Phillips, *On linear transformations*, Trans. Amer. Math. Soc. vol. 48 (1940) pp. 516–541.

Theorem 3.1 now follows from 2.29. If we take σ of that theorem to be the point 1 then $E=E_1[E]=E_\sigma[T]$ where σ consists of all $\lambda\in\sigma(T)$ with $f(\lambda)=1$.

3.3. LEMMA. *If $f_n\in\mathcal{F}(T)$ and if λ is a pole of order ν for $R_\xi(T)$ then the weak convergence of $f_n(T)$ implies the existence of the limits*

$$\lim_n f_n^{(j)}(\lambda), \qquad j=0,\cdots,\nu-1.$$

Furthermore if $f_n(T)\to f(T)$ weakly where $f\in\mathcal{F}(T)$ then

$$\lim_n f_n^{(j)}(\lambda)=f^{(j)}(\lambda), \qquad j=0,\cdots,\nu-1.$$

By 2.17, λ has index ν and so the lemma follows from 2.21 by the same argument used to derive 1.11 and 1.12 from the formula in 1.8.

3.4. THEOREM. *Let P be a polynomial not identically zero and let $f_n\in\mathcal{F}(T)$, $n=1,2,\cdots$. If $f_n(T)$ converges to a projection E with $E\mathfrak{X}\subset\mathfrak{M}[P]$ then either there are no $\lambda\in\sigma(T)$ where $f_n(\lambda)\to 1$, in which case $E=0$, or else:*

(1) *The set σ of points $\lambda\in\sigma(T)$ where $f_n(\lambda)\to 1$ consists of a finite number of poles $\lambda_1,\cdots,\lambda_k$ of $R_\xi(T)$.*

(2) *If $\nu_1,\cdots,\nu_k$ are the orders of the poles $\lambda_1,\cdots,\lambda_k$ then*

$$f_n(\lambda_i)\to 1, \qquad f_n^{(j)}(\lambda_i)\to 0, \qquad j=1,\cdots,\nu_i-1;\ i=1,\cdots,k.$$

(3) *For $i=1,\cdots,k$, λ_i is a root of $P(\lambda)$ and its multiplicity is at least ν_i.*

(4) $\prod_{i=1}^{k}(\lambda_i I-T)^{\nu_i}E=0$.

If there are no $\lambda\in\sigma(T)$ where $f_n(\lambda)\to 1$ then $E=0$ by 3.1. Now suppose the set σ of all $\lambda\in\sigma(T)$ where $f_n(\lambda)\to 1$ is not empty. We have, by 3.1, σ a spectral set of T and $E=E_\sigma[T]$. Since $E\mathfrak{X}\subset\mathfrak{M}[P]$ we have $P(T)E_\sigma(T)=0$ and so (1), (3) and (4) follow from 2.19. Statement (2) follows from the preceding lemma.

3.5. LEMMA. *Let $f_n\in\mathcal{F}(T)$, $n=1,2,\cdots$, and let $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ be a polynomial whose distinct roots are $\lambda_1,\cdots,\lambda_k$. Then if $f_n(T)P(T)\to 0$ weakly and $\lim_n f_n(\lambda_i)=\alpha_i\neq 0$, $i=1,\cdots,k$, we have*

$$\mathfrak{M}_{\lambda_i}^{\nu_i}=\mathfrak{M}_{\lambda_i}^{n}, \qquad \overline{\mathfrak{N}}_{\lambda_i}^{\nu_i}=\overline{\mathfrak{N}}_{\lambda_i}^{n}, \qquad i=1,\cdots,k;\ n\geqq\nu_i.$$

If $\lambda_i\in\rho(T)$ the conclusion is obvious. If $\lambda_i\in\sigma(T)$ then $f_n(\lambda)$ is regular at λ_i and we may expand $P(\lambda)f_n(\lambda)$ about $\lambda=\lambda_i$,

$$P(\lambda)f_n(\lambda)=\sum_{m=0}^{\nu_i}\frac{(\lambda-\lambda_i)^m}{m!}\left[P(\lambda)\cdot f_n(\lambda)\right]_{\lambda=\lambda_i}^{(m)}+(\lambda-\lambda_i)^{\nu_i+1}g_{n,i}(\lambda).$$

If the function $g_{n,i}$ which is regular in a neighborhood of λ_i is defined by the

above formula on the whole of the T-admissible domain D_n upon which f_n is defined and regular, then $g_{n,i}\in\mathcal{J}(T)$. Now since $P^{(r)}(\lambda_i)=0$, $r<\nu_i$, and $P^{(\nu_i)}(\lambda_i)\neq 0$ we have

$$[P(\lambda)f_n(\lambda)]^{(m)}_{\lambda=\lambda_i}=\begin{cases}\sum_{r=0}^{m} {}_mC_r f_n^{(m-r)}(\lambda_i)P^{(r)}(\lambda_i)=0, & m<\nu_i,\\ f_n(\lambda_i)P^{(\nu_i)}(\lambda_i), & m=\nu_i,\end{cases}$$

and so

$$(*)\qquad P(T)f_n(T)=\frac{(T-\lambda_iI)^{\nu_i}}{\nu_i!}f_n(\lambda_i)P^{(\nu_i)}(\lambda_i)+(T-\lambda_iI)^{\nu_i+1}g_{n,i}(T).$$

Since $P(T)f_n(T)\to 0$ weakly, this equation shows that for $x\in\mathfrak{M}_{\lambda_i}^{\nu_i+1}$ we have

$$\alpha_iP^{(\nu_i)}(\lambda_i)x^*(T-\lambda_iI)^{\nu_i}x=0, \qquad x^*\in\mathfrak{X}^*,$$

and since $\alpha_iP^{(\nu_i)}(\lambda_i)\neq 0$ we have $x\in\mathfrak{M}_{\lambda_i}^{\nu_i}$. Thus $\mathfrak{M}_{\lambda_i}^{\nu_i}=\mathfrak{M}_{\lambda_i}^{\nu_i+1}=\mathfrak{M}_{\lambda_i}^{n}$, $n\geqq\nu_i$. Since $P(T)f_n(T)\to 0$ weakly, (*) gives

$$\nu_i!(T-\lambda_iI)^{\nu_i+1}g_{n,i}(T)\to-\alpha_iP^{(\nu_i)}(\lambda_i)(T-\lambda_iI)^{\nu_i}\text{ weakly},$$

and since $\alpha_iP^{(\nu_i)}(\lambda_i)\neq 0$ we see that every $x\in\mathfrak{N}_{\lambda_i}^{\nu_i}$ is the weak limit of a sequence $x_n\in\mathfrak{N}_{\lambda_i}^{\nu_i+1}$. Thus $\mathfrak{N}_{\lambda_i}^{\nu_i}\subset\overline{\mathfrak{N}}_{\lambda_i}^{\nu_i+1}$ and hence

$$\mathfrak{N}_{\lambda_i}^{\nu_i+1}\subset\mathfrak{N}_{\lambda_i}^{\nu_i}\subset\overline{\mathfrak{N}}_{\lambda_i}^{\nu_i+1},\qquad \overline{\mathfrak{N}}_{\lambda_i}^{\nu_i+1}=\overline{\mathfrak{N}}_{\lambda_i}^{\nu_i}.$$

Now since $(\lambda_iI-T)P(T)f_n(T)\to 0$ weakly we conclude, by what has just been proved, that $\overline{\mathfrak{N}}_{\lambda_i}^{\nu_i+2}=\overline{\mathfrak{N}}_{\lambda_i}^{\nu_i+1}$. A repetition of this argument shows that $\overline{\mathfrak{N}}_{\lambda_i}^{n}=\overline{\mathfrak{N}}_{\lambda_i}^{\nu_i}$ for all $n\geqq\nu_i$.

3.6. THEOREM. *Let $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ be a polynomial whose distinct roots are $\lambda_1,\cdots,\lambda_k$. Let $f_n\in\mathcal{J}(T)$ satisfy*

(1) $f_n(\lambda_i)\to 1$, $f_n^{(j)}(\lambda_i)\to 0$, $i=1,\cdots,k$; $j=1,\cdots,\nu_i-1$,

(2) $P(T)f_n(T)\to 0$.

Then the following statements are equivalent.

(3) $f_n(T)\to E$, $E^2=E$, $E\mathfrak{X}=\mathfrak{M}[P]$.

(4) *Each λ_i, $i=1,\cdots,k$, is either in $\rho(T)$ or else a pole of $R_\xi(T)$.*

(5) *For $i=1,\cdots,k$, λ_i is either in $\rho(T)$ or else a pole of $R_\xi(T)$ of order at most ν_i.*

(6) $\mathfrak{N}[P]\oplus\mathfrak{M}[P]=\mathfrak{X}$, *$\mathfrak{N}[P]$ is closed.*

(7) $\mathfrak{N}_{\lambda_i}^{\nu_i+1}$, $i=1,\cdots,k$, *is closed.*

Furthermore when the limit E exists the set $\sigma=(\lambda_1,\cdots,\lambda_k)\sigma(T)$ is a spectral set of T and

$$E=E_\sigma[T],\qquad \mathfrak{M}[P]=\mathfrak{X}_\sigma,\qquad \mathfrak{N}[P]=\mathfrak{X}_{\sigma'}.$$

In view of 3.4 statement (3) implies (4) and (4) in view of 3.5 and 2.17 implies (5). From 2.23 it is seen that (5) and (6) are equivalent. Also 2.17

shows that (5) implies (7). Now assume (7). Lemma 3.5 gives

$$\mathfrak{N}_{\lambda_i}^{\nu_i+1} = \overline{\mathfrak{N}}_{\lambda_i}^{\nu_i} \supset \mathfrak{N}_{\lambda_i}^{\nu_i} \supset \mathfrak{N}_{\lambda_i}^{\nu_i+1},$$

so that $\mathfrak{N}_{\lambda_i}^{\nu_i} = \mathfrak{N}_{\lambda_i}^{n}$ for $n \geqq \nu_i$. Since $\mathfrak{N}_{\lambda_i}^{2\nu_i} = \mathfrak{N}_{\lambda_i}^{\nu_i}$ we have for every $x \in \mathfrak{X}$ a $z \in \mathfrak{N}_{\lambda_i}^{\nu_i}$ such that $(\lambda_i I - T)^{\nu_i} x = (\lambda_i I - T)^{\nu_i} z$, that is $x - z \in \mathfrak{M}_{\lambda_i}^{\nu_i}$ and so

(*) $$\mathfrak{X} = \mathfrak{M}_{\lambda_i}^{\nu_i} + \mathfrak{N}_{\lambda_i}^{\nu_i} \text{ and } \mathfrak{N}_{\lambda_i}^{\nu_i} \text{ is closed.}$$

Now if $\lambda_i \in \sigma(T)$, f_n is regular at λ_i and so for λ near λ_i

$$f_n(\lambda) = \sum_{j=0}^{\nu_i-1} \frac{f_n^{(j)}(\lambda_i)}{j!} (\lambda - \lambda_i)^j + (\lambda - \lambda_i)^{\nu_i} g_{n,i}(\lambda).$$

If $g_{n,i}(\lambda)$ is defined by the above equation for all λ in the domain D_n where f_n is defined and regular then $g_{n,i} \in \mathcal{J}(T)$ and

$$f_n(T) = \sum_{j=0}^{\nu_i-1} \frac{f_n^{(j)}(\lambda_i)}{j!} (T - \lambda_i I)^j + (T - \lambda_i I)^{\nu_i} g_{n,i}(T).$$

Thus for $x \in \mathfrak{M}_{\lambda_i}^{\nu_i}$ we have, using (1),

$$f_n(T)x = \sum_{j=0}^{\nu_i-1} \frac{f_n^{(j)}(\lambda_i)}{j!} (T - \lambda_i I)^j x \to x.$$

Now if $P_i(\lambda) = P(\lambda)/(\lambda - \lambda_i)^{\nu_i}$ we see from Lemma 1.4 that $P_i(T)x \neq 0$ providing $0 \neq x \in \mathfrak{M}_{\lambda_i}^{\nu_i}$. Thus if $x = (\lambda_i - T)^{\nu_i} y \in \mathfrak{M}_{\lambda_i}^{\nu_i}$ we have $f_n(T)x \to x$, $P_i(T)f_n(T)x = P(T)f_n(T)y \to 0$ (by (2)) and so $P_i(T)x = 0$ and hence $x = 0$. This shows that $\mathfrak{N}_{\lambda_i}^{\nu_i} \cdot \mathfrak{M}_{\lambda_i}^{\nu_i} = 0$. Thus (*) gives

$$\mathfrak{X} = \mathfrak{M}_{\lambda_i}^{\nu_i} \oplus \mathfrak{N}_{\lambda_i}^{\nu_i}, \ \mathfrak{N}_{\lambda_i}^{\nu_i} \text{ is closed.}$$

Thus 2.23 applied to the function $f(\lambda) = (\lambda_i - \lambda)^{\nu_i}$ shows us that λ_i is a pole of order at most ν_i for $R_t(T)$. Hence (7) implies (5). We have now shown the equivalence of (4), (5), (6), (7) and also the fact that (3) implies any one of them. We shall now show that these conditions imply (3). Assuming (5) we have by 2.21

$$f_n(T) = f_n(T)E_{\sigma'} + \sum_{\lambda_i \in \sigma} \sum_{j=1}^{m_i-1} \frac{(T - \lambda_i I)^j}{j!} f_n^{(j)}(\lambda_i) E_{\lambda_i}[T],$$

where $\sigma = (\lambda_1, \cdots, \lambda_k)\sigma(T)$ and m_i is the order of $\lambda_i \in \sigma$ as a pole of $R_t(T)$. Now in $\mathfrak{X}_{\sigma'}$ the spectrum of T is σ' (2.12) and hence since $0 \notin P(\sigma')$ we see that $P(T)$ is a 1-1 bicontinuous map of $\mathfrak{X}_{\sigma'}$ into all of itself. Thus $f_n(T)E_{\sigma'} = f_n(T)P(T)P^{-1}(T)E_{\sigma'} \to 0$ (by (2)). The above formula for $f_n(T)$ now shows (using (1)) that $f_n(T) \to \sum_{\lambda_i \in \sigma} E_{\lambda_i}[T] = E_\sigma[T]$. The final statements of the theorem follow from 2.23.

3.7. THEOREM. *Let $P(\lambda)$, $f_n(\lambda)$ be as in the preceding theorem and satisfy* (1) *and* (2) *of that theorem. For $i=1, \cdots, k$ let $(\lambda_i I-T)$ take bounded closed sets into closed sets. Then statements* (3), $\cdots$, (7) *of the preceding theorem are all true.*

Since $(\lambda_i I-T)$ takes bounded closed sets into closed sets it may be proved (see Banach[12] where the corresponding statement is proved for a compact T) that $(\lambda_i I-T)\mathfrak{X}$ is closed. Since $(\lambda_i I-T)^2\mathfrak{X}\subset(\lambda_i I-T)\mathfrak{X}$ a repetition of this principle gives the fact that $(\lambda_i I-T)^2\mathfrak{X}$ is closed, and so on, that is, $\mathfrak{N}_{\lambda_i}^n$ is closed for all $n=1, 2, \cdots$, and thus (7) of 3.6 is satisfied.

3.8. THEOREM. *Let $P(\lambda)$, $f_n(\lambda)$ be as in* 3.6 *and satisfy* (1), (2) *of that theorem. Suppose that* 0 *is not a root of $P(\lambda)$ and that T^n is compact for some integer n. Then* (3), $\cdots$, (7) *of* 3.6 *are all true.*

By 2.34 all spectral points $\lambda\neq 0$ of T are poles of $R_\lambda(T)$ and so 3.8 follows from 3.6. Theorem 3.8 is also an immediate corollary of 3.7.

3.9. THEOREM. *Let $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ be a polynomial whose distinct roots are $\lambda_1, \cdots, \lambda_k$. Let $f_n\in\mathcal{F}(T)$ satisfy the following:*

(1) *$f_n(\lambda_i)\to 1$, $f_n^{(j)}(\lambda_i)\to 0$, $i=1, \cdots, k$; $j=1, \cdots, \nu_i-1$.*

(2) *$P(T)f_n(T)\to 0$ strongly.*

Then the following conditions are equivalent.

(3) *$f_n(T)\to E$ strongly, $E^2=E$, $E\mathfrak{X}=\mathfrak{M}[P]$.*

(4) *$f_n(T)x$ is weakly compact, $x\in\mathfrak{X}$.*

(5) *$\mathfrak{N}[P]\oplus\mathfrak{M}[P]=\mathfrak{X}$, $|f_n(T)|\leqq M$.*

As in the proof of 3.6 we have for $y\in\mathfrak{M}_{\lambda_i}^{\nu_i}$

$$f_n(T)y=\sum_{j=0}^{\nu_i-1}\frac{f_n^{(j)}(\lambda_i)}{j!}(T-\lambda_i I)^j y\to y,$$

and so 1.4 gives the fact that $f_n(T)y\to y$ for $y\in\mathfrak{M}[P]$. Also (2) shows that $f_n(T)y\to 0$ for $y\in\mathfrak{N}[P]$. Now if we assume (4) we see that $\left|f_n(T)\right|\leqq M$ and hence $f_n(T)y\to 0$ for $y\in\overline{\mathfrak{N}}[P]$. Thus $\overline{\mathfrak{N}}[P]\cdot\mathfrak{M}[P]=0$. Since $\left|f_n(T)\right|\leqq M$ it is seen that $f_n(T)x$ converges for every x in the closure of $\mathfrak{N}[P]+\mathfrak{M}[P]$. It will now be shown that

$$(*)\qquad \overline{\mathfrak{N}}[P]\oplus\mathfrak{M}[P]=\mathfrak{X}.$$

If this is not true there is an $x\in X$ and an $x^*\in X^*$ with $x^*x=1$, $x^*\mathfrak{M}[P]=0=x^*\mathfrak{N}[P]$. By 1.4 there are points $x_i\in\mathfrak{M}[P_i]$ with $x=x_1+\cdots+x_k$ and, since $x^*x=1$, not all x^*x_i are zero. Suppose $x^*x_1\neq 0$ and pick a sequence n_i of integers so that $f_{n_i}(T)x_1\to y_1$ weakly. Since $x_1\in\mathfrak{N}[P_1]$ it is of the form $x_1=P_1(T)z_1$ and $(\lambda_1 I-T)^{\nu_1}x_1=P(T)z_1$. Thus $(\lambda_1 I-T)^{\nu_1}f_n(T)x_1=f_n(T)P(T)z_1$

[12] S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932; in particular Théoreme 11, p. 151.

$\to 0$ which shows that $(\lambda_1 I - T)^{r_1} y_1 = 0$, that is, $y_1 \in \mathfrak{M}_{\lambda_1}^{r_1} \subset \mathfrak{M}[P]$. Since $x^* \mathfrak{M}[P] = 0$ we have therefore

$$(**) \qquad x^* f_{n_i}(T) x_1 \to 0.$$

Now it is clear that it is no loss of generality to assume that $\lambda_1 \in \sigma(T)$, for if $\lambda_1 \in \rho(T)$ hypothesis (2) remains valid if $P(T)$ is replaced by $P_1(T)$ and also $\mathfrak{M}[P] = \mathfrak{M}[P_1]$. Thus we shall assume $\lambda_1 \in \sigma(T)$ and hence $f_n(\lambda)$ is regular at $\lambda = \lambda_1$. As in the proof of 3.5 we have

$$f_n(T) = \sum_{j=0}^{r_1 - 1} \frac{(T - \lambda_1 I)^j}{j!} f_n^{(j)}(\lambda_1) + (T - \lambda_1 I)^{r_1} g_n(T)$$

and hence, since $(T - \lambda_1 I)^{r_1} x_1 = (-1)^{r_1} P(T) x_1$ and $x^* \mathfrak{N}[P] = 0$, we have from (1)

$$x^* f_n(T) x_1 = x^* \sum_{j=0}^{r_1 - 1} \frac{f_n^{(j)}(\lambda_1)}{j!} (T - \lambda_1 I)^j x_1 \to x^* x_1 \neq 0.$$

The fact contradicts (**) and proves (*). Thus on the basis of (4) we have shown that

$$Ex = \lim_n f_n(T) x, \qquad x \in \mathfrak{X},$$

exists. Statement (2) shows that $P(T)E = 0$ and hence $E\mathfrak{X} \subset \mathfrak{M}[P]$. But, as was shown above, $f_n(T)x \to x$ for $x \in \mathfrak{M}[P]$ and hence $E^2 x = \lim_n f_n(T) Ex = Ex$ so that $E^2 = E$, $E\mathfrak{X} = \mathfrak{M}[P]$. Thus we have shown that (4) implies (3) and (5). It remains to be shown that (5) implies (3). But, as was shown at the beginning of the proof, (1), (2) and $|f_n(T)| \leqq M$ imply the convergence of $f_n(T)x$ for every $x \in \overline{\mathfrak{N}}[P] + \mathfrak{M}[P]$ and hence $Ex = \lim_n f_n(T)x$ exists for every $x \in \mathfrak{X}$. The above argument shows that E is a projection of the whole space onto the manifold $\mathfrak{M}[P]$. This completes the proof of the theorem.

As Theorem 3.4 shows, the condition (1) of 3.6, that is, $f_n(\lambda_i) \to 1$, $f_n^{(j)}(\lambda_i) \to 0$, $1 \leqq j \leqq \nu_i - 1$, is a necessary condition for the uniform convergence $f_n(T) \to E$, $E^2 = E$, $E\mathfrak{X} = \mathfrak{M}[P]$ providing $P(\lambda)$ is the polynomial of smallest degree for which $P(T) f_n(T) \to 0$. As the following theorem shows, a completely analogous situation holds in the case of strong convergence.

3.10. THEOREM. *Let $P(\lambda)$ be a polynomial and $f_n \in \mathcal{F}(T)$ be such that*

(1) *$f_n(T) \to E$ strongly, $E^2 = E$, $E\mathfrak{X} = \mathfrak{M}[P]$.*

Let $P'(\lambda)$ be the polynomial of smallest degree such that $\mathfrak{M}[P] = \mathfrak{M}[P']$. Let $P''(\lambda)$ be the polynomial of smallest degree such that $P''(T) f_n(T) \to 0$ strongly. Then

(2) *$P'(\lambda) \equiv P''(\lambda)$.*

(3) *If λ is a root of $P'(\lambda)$ of multiplicity ν then λ is in the point spectrum of T and has index ν and*

$$f_n(\lambda) \to 1, \qquad f_n^{(j)}(\lambda) \to 0, \qquad j = 1, \cdots, \nu - 1.$$

Let $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$, $P'(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i'}$, $P''(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i''}$. Clearly $0\leqq\nu_i'\leqq\nu_i$ and since (1) implies $P(T)f_n(T)\to 0$ strongly we see that $0\leqq\nu_i''\leqq\nu_i$. Also since, in the sense of strong convergence, we have $P'(T)f_n(T)\to P'(T)E=0$ we see that $0\leqq\nu_i''\leqq\nu_i'\leqq\nu_i$. Let μ_i be the index of λ_i. By 1.4 we have $\mathfrak{M}_{\lambda_1}^{\nu_1}\oplus\cdots\oplus\mathfrak{M}_{\lambda_k}^{\nu_k}=\mathfrak{M}_{\lambda_1}^{\nu_1'}\oplus\cdots\oplus\mathfrak{M}_{\lambda_k}^{\nu_k'}$ from which it is clear that every root λ_i of $P'(\lambda)$ has index $\mu_i\geqq\nu_i'$ and hence every root of $P'(\lambda)$ is in the point spectrum. Since f_n is regular at such a root we may write

$$f_n(T)=\sum_{j=0}^{\nu_i'-1}\frac{f_n^{(j)}(\lambda_i)}{j!}(T-\lambda_iI)^j+(T-\lambda_iI)^{\nu_i'}g_{n,i}(T).$$

If $x\neq 0$ and $(\lambda_iI-T)x=0$ we have $f_n(T)x=f_n(\lambda_i)x$ and since $x\in\mathfrak{M}_{\lambda_i}^{\nu_i'}\subset\mathfrak{M}[P]$ we have $f_n(T)x\to x$, that is $f_n(\lambda_i)\to 1$. If $\nu_i'>1$ then there is an x with $(T-\lambda_iI)x\neq 0$, $(T-\lambda_iI)^2x=0$ so that $x\in\mathfrak{M}[P]$ and

$$f_n(T)x=f_n(\lambda_i)x+f_n^{(1)}(\lambda_i)(T-\lambda_iI)x,$$

which shows that $f_n^{(1)}(\lambda_i)\to 0$. This argument may be repeated to give

$$f_n^{(j)}(\lambda_i)\to 0, \qquad j=1,\cdots,\nu_i'-1.$$

Since $0\leqq\nu_i''\leqq\nu_i'$ we see therefore that (1), (2), and (4) of Theorem 3.9 are satisfied with $P(\lambda)$ replaced by $P''(\lambda)$ and hence, by that theorem, $f_n(T)$ converges strongly to a projection of $\mathfrak{X}$ on $\mathfrak{M}[P'']$. This shows that $\mathfrak{M}[P'']=\mathfrak{M}[P]$, which proves that $\nu_i'\leqq\nu_i''$ and hence $P'(\lambda)\equiv P''(\lambda)$. It remains to be shown that the index μ_i of a root λ_i of P' is equal to its multiplicity ν_i'. We have already seen that $\mu_i\geqq\nu_i'$, and 3.5 applied to the polynomial $P''(\lambda)=P'(\lambda)$ gives $\mu_i\leqq\nu_i'$. This completes the proof of the theorem.

The discussion of weak convergence is almost identical with strong convergence, and so we shall not discuss it in full but merely state the following two theorems. These theorems may be proved by using the methods of proof employed in 3.9 and 3.10.

3.11. THEOREM. *Let $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ be a polynomial whose distinct roots are $\lambda_1,\cdots,\lambda_k$. Let $f_n\in\mathcal{F}(T)$ satisfy*

(1) $f_n(\lambda_i)\to 1$, $f_n^{(j)}(\lambda_i)\to 0$, $i=1,\cdots,k$; $j=1,\cdots,\nu_i-1$.

(2) $P(T)f_n(T)\to 0$ *weakly.*

Then the following assertions are equivalent.

(3) $f_n(T)\to E$ *weakly,* $E^2=E$, $E\mathfrak{X}=\mathfrak{M}[P]$.

(4) $f_n(T)x$ *is weakly compact,* $x\in\mathfrak{X}$.

(5) $\overline{\mathfrak{N}}[P]\oplus\mathfrak{M}[P]=\mathfrak{X}$, $|f_n(T)|\leqq M$.

3.12. THEOREM. *Let $P(\lambda)$ be a polynomial and $f_n\in\mathcal{F}(T)$ be such that*

(1) $f_n(T)\to E$ *weakly,* $E^2=E$, $E\mathfrak{X}=\mathfrak{M}[P]$.

Let $P'(\lambda)$ be the polynomial of smallest degree such that $\mathfrak{M}[P]=\mathfrak{M}[P']$. Let

$P''(\lambda)$ be the polynomial of smallest degree such that $P''(T)f_n(T)\to 0$ weakly. Then

(2) $P'(\lambda)\equiv P''(\lambda)$.

(3) *If λ is a root of $P'(\lambda)$ of multiplicity ν then λ is in the point spectrum of T and has index ν and*

$$f_n(\lambda)\to 1, \qquad f_n^{(j)}(\lambda)\to 0, \qquad j=1,\cdots,\nu-1.$$

3.13. THEOREM. *Let $P(\lambda)=\prod_{i=1}^{k}(\lambda_i-\lambda)^{\nu_i}$ be a polynomial whose distinct roots are $\lambda_1,\cdots,\lambda_k$. Let $f_n\in\mathcal{F}(T)$ be such that*

(1) $f_n(\lambda_i)\to 1, f_n^{(j)}(\lambda_i)\to 0, i=1,\cdots,k; j=1,\cdots,\nu_i-1$.

(2) $f_n(T)\to E$ *weakly,* $E^2=E$, $E\mathfrak{X}=\mathfrak{M}[P]$.

Then

(3) $f_n(T)\to E$ *strongly if and only if* $P(T)f_n(T)\to 0$ *strongly.*

(4) $f_n(T)\to E$ *uniformly if and only if* $P(T)f_n(T)\to 0$ *uniformly and* $\mathfrak{N}[P]$ *is closed.*

This is an immediate consequence of Theorems 3.6, 3.9 and 3.11.

In the following discussion of almost everywhere convergence we suppose that $\mathfrak{X}$ is composed of measurable scalar functions $x(t)$ defined for t in a set S. The functions $x\in\mathfrak{X}$ are measurable with respect to a completely additive non-negative set function Φ which is defined on a Borel field containing S. We suppose also that S is a denumerable union of sets of finite measure and that $x=0$ if and only if $x(t)=0$ almost everywhere. Addition and scalar multiplication of the functions $x(t)$ correspond to the similar operations in $\mathfrak{X}$. It is also assumed that if $x_n\to x$ in $\mathfrak{X}$ and $x_n(t)\to y(t)$ almost everywhere then $x(t)=y(t)$ almost everywhere. For a linear operator U in $\mathfrak{X}$ we shall write $U(x,t)$ for the value of the function Ux at the point t.

The following theorem of Banach[13] is fundamental.

3.14. THEOREM. *Let U_n be a sequence of linear operators in $\mathfrak{X}$ satisfying*

(1) *For every* $x\in\mathfrak{X}$, $\limsup_n |U_n(x,t)|<\infty$ *almost everywhere,*

(2) *For every x in a fundamental set in $\mathfrak{X}$ the* $\lim U_n(x,t)$ *exists almost everywhere.*

Then

(3) *For every* $x\in\mathfrak{X}$ *the* $\lim_n U_n(x,t)$ *exists almost everywhere.*

3.15. THEOREM. *Let P, f_n, E be as in 3.9 and satisfy* (1), (2), (4) *of that theorem. Then*

(1) *for every* $x\in\mathfrak{X}$, $\lim_n f_n(T)(x,t)=E(x,t)$ *almost everywhere if and only if*

---

[13] S. Banach, *Sur la convergence presque partout des fonctionelles linéaires*, Bull. Sci. Math. (2) vol. 50 (1926) pp. 36–43. This theorem of Banach has also been used by Yosida in generalizing the ergodic theorem of G. D. Birkhoff, see K. Yosida, *Ergodic theorems of the Birkhoff-Khintchine's type*, Jap. J. Math. vol. 17 (1940) pp. 31–36.

(2) *for every* $x\in\mathfrak{X}$, $\limsup_n |f_n(T)(x, t)| < \infty$ *almost everywhere, and*

(3) *for every* $x\in\mathfrak{X}$, $\lim_n P(T)f_n(T)(x, t)$ *exists almost everywhere.*

Statements (2) and (3) are clearly necessary for (1). Now conversely (3) shows that $f_n(T)(x, t)$ converges almost everywhere for $x\in\mathfrak{N}[P]$, and for $x\in\mathfrak{M}_{\lambda_i}^{\nu_i}$ we have

$$f_n(T)x = \sum_{j=0}^{\nu_i-1} \frac{f_n^{(j)}(\lambda_i)}{j!}(T-\lambda_i I)^j x,$$

and so $f_n(T)(x, t)$ converges almost everywhere for $x\in\mathfrak{M}[P]=\sum_{i=1}^{k}\mathfrak{M}_{\lambda_i}^{\nu_i}$. Thus 3.9(5) and 3.14 give the desired conclusion.

The ergodic theorem

The following theorems are obtained from the preceding ones by taking the polynomial $P(T)$ to be $I-T$, that is, we are here concerned with convergence to a projection of the whole space on the fixed points of T. For brevity we shall write $\mathfrak{M}$, $\mathfrak{N}$ in place of $\mathfrak{M}[P]$, $\mathfrak{N}[P]$ so that $\mathfrak{M}$ is the manifold of fixed points and $\mathfrak{N}=(I-T)\mathfrak{X}$. While the ergodic theorem is concerned with the convergence of $f_n(T)$ where $f_n(\lambda)=n^{-1}\sum_{r=0}^{n-1}\lambda^r$ we shall word the theorems for any sequence of functions $f_n\in\mathcal{F}(T)$ with $f_n(1)\to 1$. It should be noted that the condition $(I-T)f_n(T)\to 0$ in any of the various meanings of convergence reduces, for the particular sequence of averages mentioned above, to the condition $T^n/n\to 0$.

3.16. Theorem. *Let* $f_n\in\mathcal{F}(T)$ *satisfy* $f_n(1)\to 1$ *and* $(I-T)f_n(T)\to 0$. *Then the following statements are equivalent.*

(1) $f_n(T)\to E$, $E^2=E$, $E\mathfrak{X}=\mathfrak{M}$.

(2) *The point* $\lambda=1$ *is either in* $\rho(T)$ *or else a pole of* $R_\xi(T)$.

(3) *The point* $\lambda=1$ *is either in* $\rho(T)$ *or else a simple pole of* $R_\xi(T)$.

(4) $\mathfrak{N}\oplus\mathfrak{M}=\mathfrak{X}$, $\mathfrak{N}$ *is closed.*

(5) $(I-T)^2\mathfrak{X}$ *is closed.*

This[14] is a corollary of 3.6.

3.17. Theorem. *Let* $f_n\in\mathcal{F}(T)$ *satisfy* $f_n(1)\to 1$, $(I-T)f_n(T)\to 0$, *and let* $I-T$ *take bounded closed sets into closed sets. Then statements* (1), $\cdots$, (5) *of* 3.16 *are all true.*

This is a corollary of 3.7.

3.18. Theorem. *Let* $f_n\in\mathcal{F}(T)$ *satisfy* $f_n(1)\to 1$, $(I-T)f_n(T)\to 0$ *and let* T^n

[14] In the original statement of this theorem the assertion was made that if $\lambda=1$ is isolated in $\sigma(T)$ and if $|T^n|\leq M$ then $\lambda=1$ is necessarily a simple pole of $R_\xi(T)$. We are indebted to Professor Einar Hille for detecting an error in the attempted proof of this statement.

be compact for some integer n. Then the statements (1), · · · , (5) *of* 3.16 *are all true.*

This is a corollary of 3.8.

3.19. THEOREM. *Let* $f_n \in \mathcal{F}(T)$ *satisfy* $f_n(1) \to 1$ *and* $(I-T)f_n(T) \to 0$ *strongly. Then the following conditions are equivalent.*

(1) $f_n(T) \to E$ *strongly,* $E^2 = E$, $E\mathfrak{X} = \mathfrak{M}$.

(2) $f_n(T)x$ *is weakly compact,* $x \in \mathfrak{X}$.

(3) $\overline{\mathfrak{N}} \oplus \mathfrak{M} = \mathfrak{X}$, $|f_n(T)| \leqq \mathfrak{M}$.

This follows from 3.9.

Similar theorems concerning weak convergence and almost everywhere convergence may be obtained from 3.10, · · · , 3.15, and these are left to the reader.

It should be mentioned however in connection with the ergodic theorem of G. D. Birkhoff([15]), which is concerned with the almost everywhere convergence of $n^{-1}\sum_{r=0}^{n-1} x(\phi^r t)$, that the condition (3) of 3.15 is redundant. That is, even with no assumption concerning the measure preserving character of the point map ϕ other than the assertion that the strong (that is mean) theorem holds, we can say that $n^{-1}\sum_{r=0}^{n-1} x(\phi^r t)$ converges almost everywhere if and only if

$$(*) \qquad \limsup_n \left| \frac{1}{n} \sum_{r=0}^{n-1} x(\phi^r t) \right| < \infty \text{ almost everywhere.}$$

To see this note that (*) implies

$$(**) \qquad \limsup_n \left| x(\phi^n t)/n \right| < \infty \text{ almost everywhere.}$$

Clearly $x(\phi^n t)/n \to 0$ for a bounded function $x(t)$ and hence 3.14 shows that $x(\phi^n(t))/n \to 0$ almost everywhere for every summable function $x(t)$. Thus 3.15 shows that (*) is necessary and sufficient for

$$(***) \qquad \lim_n \frac{1}{n} \sum_{r=0}^{n-1} x(\phi^r t) = E(x, t) \text{ almost everywhere.}$$

Hence, one important problem remaining is to determine the conditions on the point map ϕ which are necessary and sufficient for (*) and hence for (***). This problem will be discussed elsewhere.

([15]) G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U. S. A. vol. 17 (1931) pp. 656–660.

YALE UNIVERSITY,
NEW HAVEN, CONN.

SETS OF UNIQUENESS AND SETS OF MULTIPLICITY

BY

R. SALEM

1.1. **Introduction.** Let E be a set of points in the interval $(0, 2\pi)$. E is said to be a set of multiplicity for trigonometrical series if there exists a trigonometrical series not vanishing identically and converging to zero outside E. If no such series exists, E is said to be a set of uniqueness. Every set of positive measure is a set of multiplicity, but there exist sets of multiplicity of measure zero. On the other hand, among perfect sets of measure zero there are sets of uniqueness as well as sets of multiplicity.

Let P be a perfect set of measure zero in $(0, 2\pi)$ and let $F(x)$ be any continuous, non-decreasing, purely singular function constructed on P, that is to say $F(x)$ is constant in every interval contiguous to P, but not everywhere.

Then it is a consequence of Riemann's classical theory of trigonometrical series that if the Fourier-Stieltjes coefficients of dF tend to zero, that is, if

$$c_n = \int_0^{2\pi} e^{nix} dF = o(1) \qquad (n \to \infty),$$

P is a set of multiplicity. But if $c_n \neq o(1)$ for every $F(x)$ of the above described type constructed on P, it is not known whether P is necessarily a set of uniqueness.

The only sets of uniqueness constructed so far are either sets of the "Hardy-Littlewood-Steinhaus type," called sets of the type H, or sets which are the sum of a finite number or of a denumerable infinity of H-sets, called sets of the type H_σ. For these results, due to Rajchman and to Nina Bary, and for the general theory we refer the reader to Zygmund [1, pp. 291–296](1).

We shall suppose henceforth that P is a perfect set of the Cantor type and of constant ratio of dissection ξ $(0<\xi<1/2)$, constructed on $(0, 2\pi)$. Such a set is obtained by dividing the fundamental interval in three parts of lengths proportional to ξ, $1-2\xi$, and ξ, respectively, and by removing the central open interval. Two intervals are left on which we perform the same operation, and so on indefinitely. (Cantor's classical ternary set corresponds to $\xi=1/3$.)

It is easy to see that any point x belonging to P is given by the formula

$$(1.1.1)\quad x = 2\pi[\epsilon_1(1-\xi) + \epsilon_2\xi(1-\xi) + \cdots + \epsilon_k\xi^{k-1}(1-\xi) + \cdots]$$

where the ϵ_i are 0 or 1. Let $F(x)$ be the continuous, non-decreasing, purely

Presented to the Society, February 27, 1943; received by the editors February 2, 1943.

(1) References to bibliography at end of this paper will appear in brackets.

singular function defined by $F(x)=\epsilon_1/2+\epsilon_2/2^2+\cdots+\epsilon_k/2^k+\cdots$ when x is given by (1.1.1). Then we have [Carleman 1, p. 225; Salem 1, p. 325]:

$$c_n=\int_0^{2\pi}e^{nix}dF=e^{\pi ni}\prod_{k=1}^{\infty}\cos\pi n\xi^{k-1}(1-\xi). \tag{1.1.2}$$

Instead of considering the Fourier-Stieltjes coefficients of dF (depending on the integral parameter n) we can consider the Fourier-Stieltjes transform of dF (depending on the continuous parameter v), F being defined to be equal to zero in $(-\infty, 0)$ and to 1 in $(2\pi, \infty)$. We have thus

$$c(v)=\int_{-\infty}^{+\infty}e^{vix}dF=e^{\pi vi}\prod_{k=1}^{\infty}\cos\pi v\xi^{k-1}(1-\xi). \tag{1.1.3}$$

If the set P and the function F are constructed on the interval $(-\pi/(1-\xi), \pi/(1-\xi))$ instead of the interval $(0, 2\pi)$, we have, denoting by Φ the function $F((1-\xi)x+\pi)$,

$$\gamma(u)=\int_{-\infty}^{+\infty}e^{uix}d\Phi=\prod_{k=0}^{\infty}\cos\pi u\xi^k. \tag{1.1.4}$$

We can interpret $\Phi(x)$ as the distribution function which is the convolution of infinitely many symmetric Bernoulli distribution functions:

$$\Phi(x)=B(x/\pi)*B(x/\pi\xi)*\cdots*B(x/\pi\xi^k)*\cdots,$$

$B(x)$ being the Bernoulli distribution function equal to zero in $(-\infty, -1)$, to $1/2$ in $(-1, +1)$ and to 1 in $(1, +\infty)$. With this interpretation, it is known (see Jessen and Wintner [1]) that ξ need no longer satisfy the condition $0<\xi<1/2$, but must satisfy the inequality $0<\xi<1$: $\gamma(u)$ is still the Fourier-Stieltjes transform of the distribution function Φ, the essential difference being that if $0<\xi<1/2$, the "spectrum" of Φ is the perfect set nowhere dense P, and Φ is purely singular; while if $1/2\leqq\xi<1$ the "spectrum" of Φ is the whole interval $(-\pi/(1-\xi), \pi/(1-\xi))$, and Φ can be either purely singular or absolutely continuous.

In what follows, c_n, $c(v)$ and $\gamma(u)$ shall always have the meaning of the formulae (1.1.2), (1.1.3) and (1.1.4), respectively.

While, if we are concerned only with the behaviour of c_n or of $\gamma(u)$ as n or u tends to ∞, we can suppose $0<\xi<1$, it is plain that in the study of perfect sets of uniqueness and of multiplicity we must suppose $0<\xi<1/2$.

1.2. It has been known for a long time that the question whether a set P of the above-described type is a set of uniqueness or a set of multiplicity is deeply related to the arithmetical properties of the number ξ (see Nina Bary [2]). Actually it has been shown by Nina Bary [3] and by Kershner [1] that if ξ is rational and equal to the irreducible fraction p/q, c_n and $\gamma(u)$ tend to zero if $p\neq 1$, and do not tend to zero if $p=1$. This, as we have seen, implies

that P is a set of multiplicity when $p \neq 1$, but not necessarily that P is a set of uniqueness when $p=1$. This latter result is, however, true, and has been proved by Nina Bary [3] who has shown that if $p=1$ the set P is of the type H, thus solving completely the question whether P is a set of uniqueness or a set of multiplicity when ξ is rational.

The purpose of the present paper is to give the solution of the same problem when ξ is irrational. This solution has been rendered possible due to an interesting work of Pisot concerning the properties of a certain set of algebraic integers which we shall define presently (see Pisot [1]—the same set has also been considered, later on, by Vijayaraghavan [1]). We shall say that α is a "Pisot-Vijayaraghavan number," or briefly a "P. V. number," if α is an algebraic integer such that all its conjugates lie inside the unit circle $|z|<1$. The relation existing between P. V. numbers and the problem of the behaviour of $\gamma(u)$ as $u \to \infty$ has first been recognized by Erdös [1], who proved that if $1/\xi$ is a P. V. number, $\gamma(u)$ does not tend to zero for $u \to \infty$. (The trivial case $\xi = 1/2$ is excluded.)

Here we shall prove the following theorems:

THEOREM I. *Let $G(x)$ be any non-decreasing function, constant in $(-\infty, 0)$ and in $(2\pi, \infty)$, and let d_n and $d(v)$ be the Fourier-Stieltjes coefficient and the Fourier-Stieltjes transform defined by*

$$d_n = \int_0^{2\pi} e^{niz} dG, \qquad d(v) = \int_{-\infty}^{+\infty} e^{viz} dG.$$

Then, if $d_n = o(1)$ for $n \to \infty$, we have also $d(v) = o(1)$ for $v \to \infty$. (The converse proposition is, of course, trivial.)

This theorem was proved first independently by A. Zygmund, whose proof, based on different ideas, will be published elsewhere.

In particular we have: *Let $0 < \xi < 1$. If $c_n = o(1)$ for $n = \infty$, then $c(v) = o(1)$ for $v = \infty$, and hence $\gamma(u) = o(1)$ for $u = \infty$.* In other words, $\gamma(u)$ and c_n both tend or do not tend to zero as u and n increase infinitely.

THEOREM II. *Let $0 < \xi < 1$. If $\gamma(u)$ does not tend to zero for $u = \infty$, then $1/\xi$ is necessarily a P. V. number.*

This theorem is the converse of the result of Erdös quoted above. Collecting the two results we see that c_n and $\gamma(u)$ tend to zero if, and only if, $1/\xi$ is not a P. V. number.

This proves that if $0 < \xi < 1/2$ and $1/\xi$ is not a P. V. number, the set P is a set of multiplicity. But it does not prove that if $1/\xi$ is a P. V. number, $0 < \xi < 1/2$, the set P is a set of uniqueness. This result is the object of Theorem III.

THEOREM III. *Let $0 < \xi < 1/2$. The set P is a set of uniqueness if (and only if)*

$1/\xi$ is a P. V. number. In this case P is a set of the type H_s.

When $\xi<1/2$ approaches $1/2$ the middle interval removed in the construction of the set P has a relative length tending to zero, and the Hausdorff dimensionality of the set, which is $\log 2/|\log \xi|$, tends to 1. It is, therefore, of some interest to prove that we can have perfect sets of uniqueness constructed with $\xi=1/(2+\epsilon)$, ϵ being positive and arbitrarily small. This is the consequence of the following theorem:

THEOREM IV. *There are P. V. numbers of the form $2+\epsilon$, ϵ being positive and arbitrarily small. Hence, there are perfect sets P of uniqueness for which the Hausdorff dimensionality is as close to 1 as we please.*

2. **Proof of Theorem I.** We have $d(v)=\int_{-\infty}^{+\infty}e^{vix}dG=\int_0^{2\pi}e^{vix}dG$ and $d_n=d(n)$. Suppose that $\lim d(n)=0$. If $d(v)$ does not tend to zero, we can find a sequence of numbers $v_1, v_2, \cdots, v_p, \cdots$ increasing infinitely and such that $|d(v_p)|>\delta$, δ being a positive number. Let $v_p=n_p+\alpha_p$, n_p being the integral part, and α_p the fractional part of v_p. There exists an infinite subsequence $\{v_q\}$ of the sequence $\{v_p\}$ such that α_q has a limit α when $q\to\infty$. We have

$$d(v_q)=\int_0^{2\pi}e^{n_q ix}e^{\alpha_q ix}dG,$$

hence

$$\left|\int_0^{2\pi}e^{n_q ix}e^{\alpha ix}dG-d(v_q)\right|<[G(2\pi)-G(0)]\max_x|e^{\alpha_q ix}-e^{\alpha ix}|<\epsilon_q,$$

ϵ_q tending to zero as $q\to\infty$. Hence, for q large enough we would have

$$\left|\int_0^{2\pi}e^{n_q ix}e^{\alpha ix}dG\right|>\delta/2.$$

But this is impossible for it is known as a consequence of Rajchman's theory of formal multiplication of trigonometrical series that $\int_0^{2\pi}e^{nix}\cos\alpha x dG$ and $\int_0^{2\pi}e^{nix}\sin\alpha x dG$ tend to zero for $n=\infty$ if, as is the case, $\int_0^{2\pi}e^{nix}dG$ tends to zero. (See Rajchman [2, p. 688].) Thus the theorem is proved.

3.1. **Proof of Theorem II.** Suppose that $0<\xi<1$ and that $\gamma(u)=\prod_{k=0}^{\infty}\cos\pi u\xi^k$ does not tend to zero for $u=\infty$. This means that there exists a sequence of numbers $u_1, u_2, \cdots, u_s, \cdots$ increasing infinitely and such that $|\gamma(u_s)|>a$, a being a positive number. Let $\theta=1/\xi$ $(1<\theta<\infty)$ and let $\theta^{n-1}<u_s\leqq\theta^n$, the integer $n=n_s$ being a function of s. We can write $u_s=\lambda_s\theta^n$ where $1/\theta<\lambda_s\leqq 1$. Hence λ_s belongs to the interval $(1/\theta, 1)$ and we can find an infinite subsequence $\{u_q\}$ of the sequence $\{u_s\}$ such that λ_q has a limit λ $(1/\theta\leqq\lambda\leqq 1)$ when $q\to\infty$. Now

$$|\gamma(u_q)|=|\gamma(\lambda_q\theta^{n_q})|\leqq|\cos\pi\lambda_q\cdot\cos\pi\lambda_q\theta\cdot\cos\pi\lambda_q\theta^2\cdots\cos\pi\lambda_q\theta^{n_q}|$$

hence

$$(1 - \sin^2 \pi\lambda_q)(1 - \sin^2 \pi\lambda_q\theta) \cdots (1 - \sin^2 \pi\lambda_q\theta^{n_q}) \geqq a^2$$

which implies

$$\exp(-\sin^2 \pi\lambda_q - \sin^2 \pi\lambda_q\theta - \cdots - \sin^2 \pi\lambda_q\theta^{n_q}) \geqq a^2,$$

that is to say

$$\sin^2 \pi\lambda_q + \sin^2 \pi\lambda_q\theta + \cdots + \sin^2 \pi\lambda_q\theta^{n_q} \leqq \log 1/a^2.$$

Let now u_r be a number belonging to the sequence $\{u_q\}$ and let $r>q$. We shall have

$$\sin^2 \pi\lambda_r + \sin^2 \pi\lambda_r\theta + \cdots + \sin^2 \pi\lambda_r\theta^{n_r} \leqq \log 1/a^2$$

and, as $r>q$, we shall have, a fortiori,

$$\sin^2 \pi\lambda_r + \sin^2 \pi\lambda_r\theta + \cdots + \sin^2 \pi\lambda_r\theta^{n_q} \leqq \log 1/a^2.$$

Let now $r\to\infty$, q being fixed. We shall have, as $\lim \lambda_r=\lambda$,

$$\sin^2 \pi\lambda + \sin^2 \pi\lambda\theta + \cdots + \sin^2 \pi\lambda\theta^{n_q} \leqq \log 1/a^2.$$

Since q is arbitrary, this implies that the series

$$\sum_0^\infty \sin^2 \pi\lambda\theta^n \tag{3.1.1}$$

converges to a sum not greater than $\log 1/a^2$. Hence:

Let $\theta=1/\xi$. If $\gamma(u)$ does not tend to zero for $u=\infty$, there exists a number λ $(1/\theta\leqq\lambda\leqq 1)$ such that the series $\sum_0^\infty \sin^2 \pi\lambda\theta^n$ converges.

Once this result is proved, Theorem II is a direct consequence of the theory of Pisot. Since, however, this theory is unnecessarily general for our particular purpose, we think it better to give here a brief account of Pisot's proof for the convenience of the reader.

3.2. Suppose that the series (3.1.1) converges. Let $w_n=\lambda\theta^n$ and let $w_n=a_n+\mu_n$ where a_n is the integer nearest to w_n and $-1/2<\mu_n\leqq 1/2$. *Then the sequence $\{a_n\}$ is a recurring sequence*, in other words

$$a_{n+k} + b_1 a_{n+k-1} + \cdots + b_k a_n = 0, \qquad n = 1, 2, 3, \cdots,$$

where k is a fixed integer and $b_1, b_2, \cdots, b_k$ are fixed rational numbers. It is known that in order to prove this it is sufficient to prove that the determinant

$$D_n = \begin{vmatrix} a_0 & a_1 & \cdots & a_n \\ a_1 & a_2 & \cdots & a_{n+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_n & a_{n+1} & \cdots & a_{2n} \end{vmatrix}$$

is equal to zero for all $n \geqq k$. k being the smallest integer having this property, k is the minimum order of the recurrence. Here we can write

$$a_{p+1} - \theta a_p = w_{p+1} - \theta w_p - (\mu_{p+1} - \theta \mu_p) = \theta \mu_p - \mu_{p+1} = \eta_p$$

hence

$$D_n = \begin{vmatrix} a_0 & a_1 - \theta a_0 & a_2 - \theta a_1 & \cdots & a_n - \theta a_{n-1} \\ a_1 & a_2 - \theta a_1 & a_3 - \theta a_2 & \cdots & a_{n+1} - \theta a_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_n & a_{n+1} - \theta a_n & a_{n+2} - \theta a_{n+1} & \cdots & a_{2n} - \theta a_{2n-1} \end{vmatrix}$$

$$= \begin{vmatrix} a_0 & \eta_0 & \eta_1 & \cdots & \eta_{n-1} \\ a_1 & \eta_1 & \eta_2 & \cdots & \eta_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_n & \eta_n & \eta_{n+1} & \cdots & \eta_{2n-1} \end{vmatrix}$$

and, by Hadamard's theorem,

$$D_n^2 \leqq \left(\sum_0^n a_p^2 \right) \left(\sum_0^n \eta_p^2 \right) \left(\sum_1^{n+1} \eta_p^2 \right) \cdots \left(\sum_{n-1}^{2n-1} \eta_p^2 \right).$$

Now the convergence of (3.1.1) is equivalent to the convergence of $\sum_0^\infty \mu_p^2$. Let $\sum_k^\infty \mu_p^2 = R_k$. We have

$$\eta_p^2 \leqq (\theta^2 + 1)(\mu_p^2 + \mu_{p+1}^2),$$

$$\sum_k^{n+k} \eta_p^2 \leqq \sum_k^\infty \eta_p^2 \leqq (\theta^2 + 1)(R_k + R_{k+1}) \leqq 2(\theta^2 + 1) R_k,$$

hence

$$D_n^2 \leqq \left(\sum_0^n a_p^2 \right) 2^n (\theta^2 + 1)^n R_0 R_1 \cdots R_{n-1}.$$

Now, as $p \to \infty$, $a_p \sim w_p = \lambda \theta^p$, hence

$$a_0^2 + a_1^2 + \cdots + a_n^2 \sim \lambda^2 (1 + \theta^2 + \cdots + \theta^{2n})$$
$$= \lambda^2 (\theta^{2n+2} - 1)/(\theta^2 - 1) < C\theta^{2n},$$

C being a constant depending on θ and λ only. Hence

$$D_n^2 \leqq C[2\theta^2(\theta^2 + 1)]^n R_0 R_1 \cdots R_{n-1};$$

as $R_n \to 0$ for $n \to \infty$, we have $D_n \to 0$ for $n \to \infty$, and as D_n is an integer there exists an integer k such that $D_{k-1} \neq 0$ and $D_n = 0$ for $n \geqq k$; and the sequence $\{a_n\}$ is a recurring sequence of minimum order k.

This, as is well known, proves that the power series $\sum_0^\infty a_n z^n$ represents a rational fraction whose denominator is $1+b_1z+\cdots+b_kz^k$. Thus

$$\sum_0^\infty a_n z^n = \frac{P(z)}{1+b_1z+\cdots+b_kz^k}, \tag{3.2.1}$$

$P(z)$ being a polynomial with rational coefficients. Now it follows from a theorem of Fatou [1, pp. 368–370, also Pólya and Szegö, 1, Aufgabe 2, pp. 142 and 357] that if a power series with integral coefficients represents a rational fraction, this fraction can be written in the form $P(z)/Q(z)$, P and Q being polynomials with integral coefficients, and $Q(0)$ being equal to 1. In other words, in (3.2.1), k being the minimum order of the recurrence, $b_1, b_2, \cdots, b_k$ are integers.

Now

$$\sum_0^\infty \mu_n z^n = \sum_0^\infty \lambda\theta^n z^n - \sum_0^\infty a_n z^n = \lambda\frac{1}{1-\theta z} - \frac{P(z)}{1+b_1z+\cdots+b_kz^k}.$$

Since $\sum_0^\infty \mu_n^2$ converges, the radius of convergence of $\sum_0^\infty \mu_n z^n$ is at least 1. Hence $1/\theta$ must be a root of $1+b_1z+\cdots+b_kz^k=0$, and the $k-1$ other roots must have moduli not less than 1. But owing to the convergence of $\sum_0^\infty \mu_n^2$, the function $\sum_0^\infty \mu_n z^n$ cannot have a pole on the circle $|z|=1$. Hence the conjugates of $1/\theta$ have their moduli *greater* than 1.

It is now sufficient to consider the equation $z^k+b_1z^{k-1}+\cdots+b_k=0$, to see that θ is an algebraic integer such that all its conjugates lie inside the unit circle $|z|<1$, that is, θ is a P. V. number, q.e.d.

(The above argument shows also immediately that λ is algebraic, but we do not need this result here.)

3.3. Conversely, if θ is a P. V. number, there exists a number λ such that the series (3.1.1) converges. Actually we can take $\lambda=1$. For if $\sigma_1, \sigma_2, \cdots, \sigma_{k-1}$ are the conjugates of θ, then $\theta^n+\sigma_1^n+\cdots+\sigma_{k-1}^n$ is an integer, and if $|\sigma_i| \leqq \delta < 1$ for $i=1, 2, \cdots, k-1$, we have

$$|\sigma_1^n| + \cdots + |\sigma_{k-1}^n| \leqq (k-1)\delta^n < \delta'^n \tag{3.3.1}$$

where $\delta'<1$, if n is large enough. Hence the series $\sum_0^\infty \sin^2 \pi\theta^n$ converges, and it converges like a geometric progression.

This requires (except for the trivial case $\theta=2$) that $\gamma(u)$ does not tend to zero for $u\to\infty$, as has been proved by Erdös [1]. The proof is immediate since the convergence of $\sum_0^\infty \sin^2 \pi\theta^n$ requires that the infinite product $\prod_{n=0}^\infty \cos^2 \pi\theta^n$ converges to a number $A\neq 0$, unless one of the elements of the product vanishes. This is impossible since it would require $\theta^n=(2k+1)/2$ for some positive integers n and k, which is in contradiction to the nature of the number θ. Hence for p large enough $\prod_0^p \cos^2 \pi\theta^n > A/2$ and

$$|\gamma(\theta^p)| = |\cos \pi\theta^p \cdot \cos \pi\theta^{p-1} \cdots \cos \pi\theta \cdot \cos \pi| \cdot |\cos \pi/\theta \cos \pi/\theta^2 \cdots)$$
$$> (A/2)^{1/2} |\cos \pi/\theta \cos \pi/\theta^2 \cdots|$$

and the last infinite product is greater than a positive constant except in the trivial case $\theta=2$. (The case $\theta^n=2$ with $n>1$ is incompatible with the nature of θ.)

4. **Proof of Theorem III.** Let now ξ be such that $0<\xi<1/2$ and let $\theta=1/\xi$. We shall prove that if θ is a P. V. number the set P defined in 1.1 is a set of the type H_0, hence a set of uniqueness. If x belongs to P, we have by (1.1.1)

$$x = 2\pi(\theta - 1)[\epsilon_1/\theta + \epsilon_2/\theta^2 + \cdots], \qquad \epsilon_i = 0 \text{ or } 1.$$

Let G be the set of the points $y=x/(\theta-1)$. This set lies in the interval $(0, 2\pi/(\theta-1)<2\pi)$ and we have $y=2\pi[\epsilon_1/\theta+\epsilon_2/\theta^2+\cdots]$.

Let h be a fixed integer, to be chosen later on. Consider the 2^h possible combinations of the h digits $d_1d_2 \cdots d_h$ where $d_i=0$ or 1. Let $C_1, C_2, \cdots, C_{2^h}$ denote these combinations. Writing

$$(4.1) \quad y = 2\pi[(\epsilon_1/\theta + \cdots + \epsilon_h/\theta^h) + (\epsilon_{h+1}/\theta^{h+1} + \cdots + \epsilon_{2h}/\theta^{2h}) + \cdots]$$

consider the sequence of the ϵ_i belonging to one group of terms in parentheses. They form one of the 2^h combinations $C_1, C_2, \cdots, C_{2^h}$.

Denote by G_s, $s=1, 2, \cdots, 2^h$, the subset of G such that if y belongs to G_s the combination C_s occurs in infinitely many groups of terms in parentheses when y is written in the form (4.1). Then obviously $G=G_1+G_2+\cdots+G_{2^h}$ (some of the G_s's may be empty). The G_s's are not necessarily closed sets. But as G is closed we can write, G_s' being the derived set of G_s,

$$G = (G_1 + G_1') + (G_2 + G_2') + \cdots + (G_{2^h} + G_{2^h}').$$

Now let us consider G_s for a fixed s. Let $m_1, m_2, \cdots, m_n, \cdots$ be the infinite sequence of the ranks of the groups of terms in parentheses where the combination C_s occurs. Let z be any point belonging to G_s. We have

$$\begin{aligned} z\theta^{hm_n} &= 2\pi[\epsilon_{hm_n} + \epsilon_{hm_n-1}\theta + \cdots + \epsilon_{hm_n-(h-1)}\theta^{h-1}] \\ &+ 2\pi[\epsilon_{h(m_n-1)}\theta^h + \cdots + \epsilon_1\theta^{hm_n-1}] \\ &+ 2\pi[\epsilon_{hm_n+1}/\theta + \epsilon_{hm_n+2}/\theta^2 + \cdots]. \end{aligned} \tag{4.2}$$

We can write

$$z\theta^{hm_n} = \alpha(s) + \beta(z) + \gamma(z) \tag{4.3}$$

where $\alpha(s)$, $\beta(z)$ and $\gamma(z)$ denote the three successive brackets in (4.2). It is essential to observe that while $\beta(z)$ and $\gamma(z)$ depend on the particular point z, the number $\alpha(s)$ depends on s only but not on z.

Putting $\theta^n = a_n + \mu_n$ where a_n is the integer nearest to θ^n and $-1/2 < \mu_n \leqq 1/2$, we have

$$\beta(z) = 2\pi[\epsilon_{\lambda(m_n-1)}a_\lambda + \cdots + \epsilon_1 a_{\lambda m_n - 1}] + 2\pi[\epsilon_{\lambda(m_n-1)}\mu_\lambda + \cdots + \epsilon_1 \mu_{\lambda m_n - 1}] = 2\pi N(z) + \beta_1(z), \tag{4.4}$$

where $N(z)$ is an integer. On the other hand, θ being a P. V. number, we have by (3.3.1) (h being supposed large enough)

$$|\beta_1(z)| < 2\pi[\delta'^\lambda + \delta'^{\lambda+1} + \cdots] = 2\pi\delta'^\lambda/(1-\delta') \tag{4.5}$$

where δ' is a fixed number, depending on θ, and essentially inferior to 1. We have also

$$z\theta^{\lambda m_n} = z a_{\lambda m_n} + z\mu_{\lambda m_n} \tag{4.6}$$

where

$$|z\mu_{\lambda m_n}| < 2\pi\delta'^{\lambda m_n} < 2\pi\delta'^\lambda. \tag{4.7}$$

Finally, let us remark that $y(z)$ is a point belonging to G_s, hence to G.

Let l be the length of the greatest interval contiguous to G, and let us suppose that h has been chosen such that the second members of (4.5) and (4.7) are both less than $l/8$, say. Then we have by (4.3), (4.4), and (4.6)

$$z a_{\lambda m_n} = 2\pi N(z) + \alpha(s) + y(z) + \beta_2(z)$$

where $\beta_2(z) = \beta_1(z) - z\mu_{\lambda m_n}$, and so $|\beta_2(z)| < l/4$. This is equivalent to

$$z a_{\lambda m_n} \equiv \alpha(s) + y(z) + \beta_2(z) \pmod{2\pi}.$$

Let us denote by G_s^n the set of points deduced from G_s by multiplication by the integer $a_{\lambda m_n}$ and reduction modulo 2π. Apart from the translation $\alpha(s)$ which is independent of z, the points of G_s^n are given by the formula $y(z)+\beta_2(z)$. As $|\beta_2(z)| < l/4$, and as $y(z)$ belongs to G (whose greatest contiguous interval is of length l) it follows that G_s^n has a contiguous interval of length not less than $l - 2l/4 = l/2$, and this is true for all n.

Now if we consider, instead of G_s, the closed set $G_s + G_s'$, the multiplication by $a_{\lambda m_n}$ and the reduction modulo 2π will give us the closed set $G_s^n + (G_s^n)'$ where $(G_s^n)'$ denotes the derived set of G_s^n. It is obvious that for every n, the set $G_s^n + (G_s^n)'$ will have also a contiguous interval of length not less than $l/2$, and this means [Rajchman 1, pp. 403–406] that the set $G_s + G_s'$ is a set of the type H, hence a closed set of uniqueness.

Now it is known [Nina Bary 1, p. 79, and Zygmund 1, p. 296] that the sum of closed sets of uniqueness is a set of uniqueness, hence the set $G = \sum_{s=1}^{2^h}(G_s + G_s')$ is a set of uniqueness: indeed, it is a set of the type H_σ.

Finally, the given set P is homothetic to G and they both lie in $(0, 2\pi)$. Hence, by a theorem of Marcinkiewicz and Zygmund [1], P is a set of unique-

ness. Moreover, it follows from a theorem of Nina Bary [3, p. 714] that P being homothetic to a set of the type H_σ is itself of the type H_σ.

This completes the proof of Theorem III.

5. **Proof of Theorem IV.** In order to prove that there exist P. V. numbers of the form $2+\epsilon$, ϵ being positive and arbitrarily small, we shall use a method inspired by the work of Vijayaraghavan [1, p. 350]. Consider the equation $f(z)=z^n-2z^{n-1}-1=0$, $n>2$. As $f(2)=-1$, there is a root greater than 2. We shall prove that for $|z|=1+\eta$ and for all $\eta>0$ small enough ($\eta<\eta_0$) we have $|z^n|<|2z^{n-1}+1|$. For let $z=\rho e^{i\omega}$. We have $|z^n|^2=\rho^{2n}$ and

$$|2z^{n-1}+1|^2 = 1 + 4\rho^{n-1}\cos(n-1)\omega + 4\rho^{2n-2} \geqq 4\rho^{2n-2} - 4\rho^{n-1} + 1,$$

hence it is sufficient to prove that

$$\phi(\rho) = \rho^{2n} - 4\rho^{2n-2} + 4\rho^{n-1} - 1 < 0.$$

But $\phi(1)=0$ and $(d\phi/d\rho)_{\rho=1}=-2n+4<0$. Hence there exists an η_0 such that if $\rho=1+\eta$ $(0<\eta<\eta_0)$ we have $|z^n|<|2z^{n-1}+1|$. This, by Rouché's theorem, proves that $f(z)=0$ has in the circle $|z|=1+\eta$, $0<\eta<\eta_0$, the same number of roots as $2z^{n-1}+1=0$, that is to say $n-1$ roots. Hence, letting $\eta\to 0$, we see that $f(z)=0$ has $n-1$ roots with moduli not greater than 1. But it is easy to see that there are no roots of modulus 1, for that would imply

$$(5.1) \qquad \begin{cases} \cos n\omega - 2\cos(n-1)\omega = 1, \\ \sin n\omega - 2\sin(n-1)\omega = 0. \end{cases}$$

Squaring and adding, we get $1-4\cos\omega+4=1$, that is to say $\omega=0$ which is incompatible with the first of the equations (5.1).

Hence the real root greater than 2 is a P. V. number. But

$$f(2(1+\epsilon)) = 2^n(1+\epsilon)^n - 2^n(1+\epsilon)^{n-1} - 1 = 2^n\epsilon(1+\epsilon)^{n-1} - 1 > 2^n\epsilon - 1,$$

hence taking $\epsilon=1/2^n$ we see that $f(2(1+1/2^n))$ is positive. Hence the P. V. number is included between 2 and $2(1+1/2^n)$ and as n is as large as we please, this proves the theorem.

6. **Remark.** In a previous paper [Salem 2, p. 71] the author has considered symmetrical perfect sets Q of order d (d being a positive integer) and of constant ratio of dissection, whose points are given by the formula

$$x = 2\pi[(\epsilon_1/d)(1-\xi) + (\epsilon_2/d)\xi(1-\xi) + \cdots + (\epsilon_k/d)\xi^{k-1}(1-\xi) + \cdots]$$

where $0<\xi<1/(d+1)$ and the ϵ_i take all values $0, 1, \cdots, d$. (The classical case considered in the present paper corresponds to $d=1$.) It has been proved (loc. cit. pp. 77–79) that if ξ is a rational fraction p/q (irreducible), Q is a set of multiplicity if $p\neq 1$, and a set of uniqueness if $p=1$.

The method used to prove this result, combined with the argument of the present paper, proves that Q is a set of uniqueness if and only if $\theta=1/\xi$ is a P. V. number ($\theta>d+1$).

BIBLIOGRAPHY

BARY, NINA
1. *Sur l'unicité du développement trigonométrique*, Fund. Math. vol. 9 (1927) pp. 62–118.
2. *Sur la nature diophantique du problème d'unicité du développement trigonométrique*, C. R. Acad. Sci. Paris vol. 202 (1936) p. 1901.
3. *Sur le rôle des lois diophantiques dans le problème d'unicité du développement trigonométrique*, Rec. Math. (Mat. Sbornik) N.S. vol. 2 (1937) pp. 699–722.

CARLEMAN, T.
1. *Sur les équations intégrales singulières a noyau réel et symétrique*, Uppsala University Årsskrift, 1923.

ERDÖS, P.
1. *On a family of symmetric Bernoulli convolutions*, Amer. J. Math. vol. 61 (1939) pp. 974–976.

FATOU, P.
1. *Séries trigonométriques et séries de Taylor*, Acta Math. vol. 30 (1936) pp. 335–400.

JESSEN, B. AND WINTNER, A.
1. *Distribution functions and the Riemann zeta function*, Trans. Amer. Math. Soc. vol. 38 (1935) pp. 48–88.

KERSHNER, R.
1. *On singular Fourier-Stieltjes transforms*, Amer. J. Math. vol. 58 (1936) pp. 450–452.

MARCINKIEWICZ, J. AND ZYGMUND, A.
1. *Two theorems on trigonometrical series*, Rec. Math. (Mat. Sbornik) N. S. vol. 2 (1937) pp. 733–737.

PISOT, C.
1. *La répartition modulo 1 et les nombres algébriques*, Annali di Pisa (2) vol. 7 (1938) pp. 205–248.

PÓLYA, G. AND SZEGÖ, G.
1. *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925.

RAJCHMAN, A.
1. *Sur la multiplication des séries trigonométriques et sur une classe remarquable d'ensembles fermés*, Math. Ann. vol. 95 (1926) pp. 389–408.
2. *Une classé de séries trigonométriques qui convergent presque partout vers zero*, Math. Ann. vol. 101 (1929) pp. 686–700.

SALEM, R.
1. *The absolute convergence of trigonometrical series*, Duke Math. J. vol. 8 (1941) pp. 317–334.
2. *On singular monotonic functions of the Cantor type*, Journal of Mathematics and Physics vol. 21 (1942) pp. 69–82.

VIJAYARAGHAVAN, T.
1. *On the fractional parts of the powers of a number*, Proc. Cambridge Philos. Soc. vol. 37 (1941) pp. 349–357.

ZYGMUND, A.
1. *Trigonometrical series*, Warszawa-Lwow, 1935.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

PROJECTIVE PLANES

BY

MARSHALL HALL

1. Introduction. The Theorem of Desargues has played a central role in the study of the foundations of projective geometry. Its validity has been shown to be equivalent to the possibility of the introduction of homogeneous coordinates from a skew-field, and also to the possibility of imbedding a plane in a space of higher dimension. A synthetic proof of the theorem, using a third dimension, is relatively easy, but two dimensional proofs failed and it was eventually realized that projective planes existed in which the theorem was not valid. Non-Desarguesian planes have received attention primarily as curious counter-examples, proving the necessity of assuming, along with the axioms of connection, either the existence of a third dimension, or axioms of congruence(1). But the full class of projective planes has a much better claim than this for attention. Projective planes are the logical basis for the investigation of combinatorial analysis, such topics as the Kirkman schoolgirl problem and the Steiner triple systems being interpretable directly as plane configurations, various problems in non-associative or nondistributive algebras, and the problems of universal identities in groups and skew-fields, as well as certain problems on structure (lattice) theory and the theory of partially ordered sets(2).

The author has made an attempt to investigate projective planes as systematically as possible. A considerable portion of these investigations is not included here either because of the incomplete and unsatisfactory nature of the results or because of lack of generality. War duties have forced the postponement of the completion of this work. The investigations here proceed along two main lines. §§2, 3, and 4 are concerned with methods of constructing the most general projective planes, and demonstrating the extent of variety which exists in the class of projective planes. In §4, the *free* planes, comparable in certain ways to free groups, are studied. Free planes, generated by finite configurations, are characterized (Theorems 4.8 and 4.13) and a

Presented to the Society, September 10, 1940; received by the editors October 22, 1942.

(1) Certainly this is Hilbert's attitude. See D. Hilbert [1]. This reference and all others in the paper are to the bibliography at the end of the paper.

(2) In a recent Colloquium publication it was said "The study of plane projective geometries takes us into one of the most mysterious and fascinating parts of combinatory analysis—a subject concerning which little is known." On the Kirkman problem, see T. P. Kirkman [1], E. W. Davis [1], and the bibliography on the problem, O. Eckenstein [1]. For triple systems, see A. Emch [1]. Geometric algebras related to quantum theory are investigated in P. Jordan, J. von Neumann, and E. Wigner [1]. F. Levi [1] applies the theory of configurations to combinatorial topology.

canonical set (Theorem 4.12) is determined. As a corollary it follows that the subplanes of a free plane are free planes. This analogy with free groups does not seem to go so far as to enable us to consider an arbitrary plane as a homomorphic image of a free plane. Certainly this cannot be done by a lattice homomorphism, and no clearly defined homomorphism of partial ordering is determined. If, as a consequence of identifications, we find two different points lying on two different lines, we must identify the points or identify the lines, but there is no clear reason for preferring one identification to the other. This fundamental difficulty has no analogue in the corresponding problem for groups.

§§5 and 6 proceed in a direction different from, but in a sense complementary to, the direction of the earlier sections. The earlier sections are concerned mainly with the geometry of *open configurations*, the later ones with the geometry of *confined configurations*. (The definitions of these two terms immediately precede and follow Theorem 4.8.) An effort is made to provide an algebraic substitute in the general plane for the coordinatizing skew-field in Desarguesian planes. Two related systems of coordinates are introduced, one a system of permutations (essentially projections), the other a ring defined by a ternary operation. It is perhaps unfortunate that the permutations force two points into a special role and that the ternary ring places a single quadrilateral in a special role. Theorems on configurations are associated with algebraic relations in these systems. The algebraic consequences of these relations are of great importance in characterizing the geometric conclusions to be drawn, and enable us to determine the geometric dependence or independence of configuration theorems. Many algebraic problems and new types of algebraic systems arise from natural geometric questions, and the fullest investigation of these is called for. The same geometric system (such as a web or a complete plane) may be associated with two or more non-isomorphic algebraic systems, each of which may be constructed from any other by certain rules. (See rule (5.12) in Theorem 5.9 and the examples in appendix II.) This geometric equivalence, stronger than algebraic isomorphism, is a valuable algebraic tool, comparable to the theory of classes of algebras.

2. **The axioms of projective geometry: The projective plane.** Hilbert[3] who was as much interested in metric geometry as in projective geometry bases his studies of the foundations of geometry on five sets of axioms: I. Axioms of connection, II. Axioms of order, III. Axioms of parallels, IV. Axioms of congruence, and V. Axioms of continuity. Of these only I and III define abstract Euclidean space and the introduction of ideal elements at infinity reduces this space to a projective space. Axioms II, IV, and particularly V restrict the coordinate skew-field (given by the von Staudt construction using I and III) to the real number field. Veblen and Young[4],

[3] D. Hilbert [1].

[4] O. Veblen and J. W. Young [1].

being concerned solely with projective geometry, proceed somewhat differently to the same end. More recently Garrett Birkhoff(5) has defined projective geometry as an irreducible complemented modular lattice (complemented Dedekind structure) and shows the equivalence of his axioms to those of Veblen and Young.

The principal result of these investigations is that the points of a projective space of three or more dimensions, which is necessarily Desarguesian, or a Desarguesian plane may be coordinatized by the von Staudt construction with elements from a skew-field, and the $(k-1)$-dimensional subspaces of a k-dimensional projective space contain those points which satisfy a linear equation. The skew-field will be a commutative field if and only if the Theorem of Pappus holds, while further conditions are required if it is wished to restrict the field to the field of real numbers.

Coming to the structure of projective spaces of less than three dimensions, the void space or a point can hardly be said to have any structure, while a line is completely determined by the number of points on it. This leaves only the non-Desarguesian planes, whose structure is non-trivial, to be characterized by algebraic means or otherwise.

Before proceeding to a study of projective planes it seems desirable to axiomatize the plane as simply as possible. The following axioms may be used:

P1. *Any two distinct points are contained in one and only one line.*

P2. *Any two distinct lines contain one and only one point in common.*

P3. *There exist four points, no three of which are contained in one line.*

The proof that these axioms define a plane in exactly the sense of Birkhoff or Veblen and Young is quite simple and will not be given here. P3 serves the purpose of axiom EO of Veblen and Young and might easily be replaced by a number of equivalent conditions. It seems more elegant to the author to require the existence of only four points of a specified kind rather than the infinity of conditions in axiom EO.

The removal of a line and the points on it from a projective plane π leaves a Euclidean plane π^* whose points and lines satisfy the following axioms:

EP1. *Any two distinct points are contained in one and only one line.*

EP2. *Two distinct lines contain at most one point in common.*

EP3. *The plane contains four points, no three of which are contained in one line.*

EP4. *Through a point P not contained in a line L there is exactly one line M which has no point in common with L.*

From EP4 two lines which are parallel to (that is, do not intersect) a third line are parallel to each other. If to π^* we add a point of intersection for

(5) Garrett Birkhoff [1]. Also K. Menger [1, 2]. These papers achieve independently virtually the same formulation.

each system of parallels and then add a single line containing all these new points we obtain a projective plane π. Clearly the Euclidean plane π^* determines π uniquely.

We may include as planes the systems satisfying the axioms P1 and P2 but not P3. These are easily found to be

1. the void plane,
2. a plane containing a single point and no lines,
3. a plane containing a single line and no points,
4. a plane containing a single point and one or more lines through the point,
5. a plane containing a single line and one or more points on the line,
6. a plane containing a single line and a single point not on the line,
7. a plane containing a point P on a line L with one or more additional lines through P and one or more additional points on L,
8. a plane containing a line L, and one or more points on L, a point P not on L and one or more lines through P joining P to the points of L.

These systems will be referred to as degenerate or improper planes. Note that every plane in which there are enough points on each line contains all these types of degenerate planes. In defining proper planes P3 may be replaced by any number of other axioms which exclude these degenerate cases.

Why is it that projective spaces of three or more dimensions are sufficiently regular to be coordinatized by skew-fields while projective non-Desarguesian planes exist which cannot be so coordinatized? One way of answering this fundamental question is the following: In a space configuration of three or more dimensions the identification of a line, constructed as the intersection of two planes, with a line, constructed as the union of two points forces the identification of further constructed elements and the establishment of non-trivial configurations such as the Desargues configuration, while in the plane a line may be constructed only as a union of two points, a point as the intersection of two lines and there are no forced identifications. Consider a proof of the Theorem of Desargues for the case in which the triangles are not coplanar. (See Fig. 1.)

Given: *Triangles* $A_1B_1C_1$, $A_2B_2C_2$, *not in the same plane, and* A_1A_2, B_1B_2, *and* C_1C_2 *passing through the point* P.

To prove: *That* A_1B_1 *and* A_2B_2, A_1C_1 *and* A_2C_2, B_1C_1 *and* B_2C_2 *intersect in points* C_3, B_3, *and* A_3 *which are collinear.*

Proof. A_1B_1 and A_2B_2 both lying in the plane PA_1B_1 must intersect in a point C_3. Similarly A_1C_1 and A_2C_2 intersect in a point B_3 and B_1C_1 and B_2C_2 intersect in a point A_3. But A_3, B_3, and C_3 must lie in both the plane of $A_1B_1C_1$ and the plane of $A_2B_2C_2$ and hence on m, the line of intersection of these two planes. Hence the lines A_3B_3, A_3C_3, and B_3C_3 must coincide with this intersection and the points A_3, B_3, and C_3 are collinear.

Here the kernel of the proof lies in the identification of A_3B_3, A_3C_3, and B_3C_3 with m and hence with each other. As we shall see in the next section there are no such forced identifications in the plane.

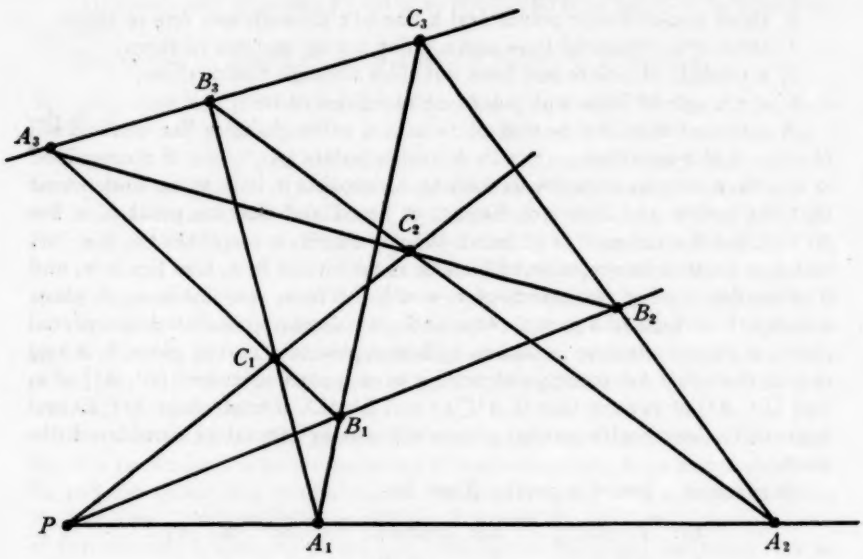


FIG. 1

3. **The generation of projective planes.** Given any set of points and lines in a plane, we may obtain further points and lines of the plane by constructing the lines joining points of the set and by finding the points of intersection of lines of the set. We may repeat this process on the enlarged set and continue as often as we please. All points and lines obtainable in this way are said to be generated by the given set. For example, all points on a line and two points not on the line generate the entire plane in four steps. But obviously the entire plane may be generated by the two points and only part of the points on the line or in various other ways. The rational plane (rational net in Veblen-Young) is generated by any quadrilateral in it.

We may use the method of generation to construct all possible projective planes.

DEFINITION. *A partial plane is any system of points and lines such that:*

p1. *There is at most one line through any two distinct points.*

p2. *There is at most one point common to two distinct lines.*

A partial plane is degenerate if it is contained in a degenerate plane. A partial plane is proper if it generates a proper plane. This it will do if and only if it contains one of the following:

1. four points, no three collinear,
2. four lines, no three copunctal,
3. three non-collinear points and a line not through any one of them,
4. three non-copunctal lines and a point not on any one of them,
5. a triangle of points and lines not sides through two of these,
6. a triangle of lines and points on two sides of it.

A *representation* of a partial plane π is a table giving a list of the lines $(L_1, \cdots)$ of π and listing after each line L_1 points $(A_1, \cdots)$ of π contained in L_1. Such a representation is said to be *explicit* if it is to be understood that the points and lines are distinct as listed and that no point of π lies on any line of π unless it is so listed. We say that π_2 is *complete* in π_1 if $\pi_1 \supset \pi_2$ and any point of intersection of lines of π_2 which lies in π_1 also lies in π_2 and if every line joining two points of π_2 which lies in π_1 also lies in π_2. A plane satisfies the axioms of a partial plane and may also be represented as a partial plane. A partial plane π_1 is said to be isomorphic to a partial plane π_2 if and only if there is a 1-1 correspondence between a representation $[L^1, A^1]$ of π_1 and $[L^2, A^2]$ of π_2 such that if $A^1 \subset L^1$ and $L^1 \rightleftarrows L^2$, $A^1 \rightleftarrows A^2$ then $A^2 \subset L^2$ and conversely. Isomorphic partial planes will not in general be considered distinct.

A representation of a partial plane is

$$
\begin{array}{lll}
L_1\colon\ PA_1A_2, & L_5\colon\ A_2B_2C_3, & L_9\colon\ B_2C_2A_3, \\
L_2\colon\ PB_1B_2, & L_6\colon\ A_1C_1B_3, & L_{10}\colon\ A_3B_3, \\
L_3\colon\ PC_1C_2, & L_7\colon\ A_2C_2B_3, & L_{11}\colon\ A_3C_3, \\
L_4\colon\ A_1B_1C_3, & L_8\colon\ B_1C_1A_3, & L_{12}\colon\ B_3C_3.
\end{array} \tag{3.1}
$$

If $P_1, \cdots, P_s$ are points no two of which are joined in the representation of a partial plane, we may extend the representation formally by adding a line $L\colon P_1, \cdots, P_s$ to the representation, or if $L_1, \cdots, L_r$ are lines no two of which intersect, we may extend by adjoining a point P to each of $L_1, \cdots, L_r$. The larger table is a representation of a partial plane which contains the original partial plane but which now contains a line L joining $P_1, \cdots, P_s$ or in which $L_1, \cdots, L_r$ now intersect in P. This method of extension may be used to imbed any partial plane in a complete plane.

THEOREM 3.1. *Given a partial plane π_1, and a representation R_1 of π_1. If π_1 is degenerate, then any plane with enough points on each line contains π_1 explicitly. If π_1 is nondegenerate, there is a complete plane π generated by π_1 in which the representation of π_1 by R_1 is explicit.*

Proof. It has already been noted above that a degenerate plane is con-

tained in any plane with enough points on each line. The representation of a degenerate plane is necessarily explicit.

Given a nondegenerate partial plane π_1 and a representation R_1 of π_1. Form sets $S_i(P^1_{i1}, \cdots, P^1_{ir})$ of points P^1 of π_1 such that:

(1) Every pair of points P^1, Q^1 of π_1 not joined by a line of π_1 is in exactly one set S.

(2) No pair of points P^1, Q^1 in a set S is joined by a line of π_1.

We may certainly define such sets in at least one way. We may take each set S with only two points in it, and include a point pair P, Q as a set if and only if P and Q are not joined in π_1. We now define a partial plane π_2 by a representation R_2 in which there are points P^1 and lines L^0, L^2. P^1 and L^0 are the points and lines of R_1 and $P^1 \subset L^0$ in R_2 if and only if $P^1 \subset L^0$ in R_1. For each set $S_i(P^1_{i1}, \cdots, P_{ir})$ we define a line L^2_i in R_2 and say that $P^1_{ij} \subset L^2_i$, $j=1, \cdots, r$, and $P^1 \not\subset L^2_i$ otherwise. Then R_2 is a representation of a partial plane π_2 containing π_1 explicitly. Every pair of points in π_1 is joined by a line in π_2 if not already joined in π_1. Similarly we may extend R_2 to a representation R_3 of a partial plane π_3 which contains π_2 explicitly and such that every pair of lines in π_2 intersects in a point of π_3 if the pair does not already intersect in a point of π_2.

We may continue this process a denumerable number of times. For each R_{2n} of a π_{2n} we form sets of two or more nonintersecting lines as was done for R_2 and use these sets to define new intersection points in a representation R_{2n+1} of a π_{2n+1} containing π_{2r} explicitly. For each R_{2n+1} of a π_{2n+1} we form sets of two or more points which are not collinear in R_{2n+1} and use these sets to define new lines in a representation R_{2n+2} containing R_{2n+1} explicitly. Every incidence of points and lines in a representation R_i will occur in a representation R_j, $j>i$, and no incidences of points and lines of R_i will occur in R_j, $j>i$, except those already in R_i. Requirement (1) of our extensions assures us that we do not adjoin two new intersection points to a pair of lines, or two new lines through a pair of points. Requirement (2) assures us that we do not adjcin a new intersection point for lines which already intersect or a new line through points already joined. Hence every R_n is a partial plane.

Now define a representation R of points and lines P, L occurring in any R_n, $n=1, 2, \cdots$, and let $P^i \subset L^j$ in R if and only if $P^i \subset L^j$ in some R_n. This determination of incidence is unambiguous, for it is determined in $R_{\max(i,j)}$, the first R_n containing both P^i and L^j, and is not altered in any succeeding R_n. Hence R contains R_1 explicitly, and satisfies the requirements for a partial plane. Every pair of lines of R_{2s} has an intersection either in R_{2s} or in R_{2s+1}, and every pair of points in R_{2s+1} is joined by either a line of R_{2s+1} or by a line of R_{2s+2}. Hence R is not only a partial plane but is a complete plane.

It is in the sense of this theorem that it was said in §2 that there are no forced identifications in projective planes. For example, (3.1) is a representa-

tion of a partial plane which may be extended to a complete plane without forcing the identification of L_{10}, L_{11}, and L_{12} as would be necessary in any Desarguesian projective space.

The consequences of this theorem are numerous when the very general nature of a partial plane is taken into account. Let us call a *subplane* π' of a plane π a set of points and lines which satisfy P1, 2, 3 or merely P1, 2. Now we may take as a partial plane any set of disjoint planes and generate a plane containing these as subplanes.

THEOREM 3.2. *An arbitrary set of planes may occur among the subplanes of a plane.*

The subplanes of a plane form a structure-lattice if we take the crosscut $\pi_1 \cap \pi_2$ of subplanes π_1, π_2 to be the points and lines common to both subplanes and the union $\pi_1 \cup \pi_2$ to be the subplane generated by π_1 and π_2. Here it is necessary to include degenerate subplanes. Unfortunately the structure-lattice of subplanes need not be particularly regular. At least it is very easy to construct a plane in which this structure-lattice is not a Dedekind structure (modular lattice). Let us take a 7-point plane π_1 and a 13-point plane π_2 with no points or lines in common and let a line of π_1 intersect a line of π_2 in a point P. Now construct a 21-point plane π_3 generated by π_1 and P and let the partial plane of π_2 and π_3 be extended to a complete plane π (necessarily larger than π_3). We now have the following diagram from the structure-lattice. Here $\pi_3 \supset \pi_1$ and $\pi_3 = \pi_3 \cap (\pi_2 \cup \pi_1) \neq (\pi_3 \cap \pi_2) \cup \pi_1 = \pi_1$, and the subplane structure-lattice is certainly not modular.

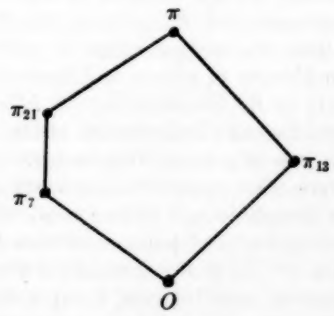


FIG. 2

The great generality of projective planes has so far been demonstrated in showing that (1) the configurations in different parts are independent and (2) the structure-lattice of subplanes is not greatly restricted, at least in any familiar way.

4. Free extensions and free planes. The simplest way of extending a par-

tial plane is to let each new point adjoined be the intersection of only two old lines and each new line join only two old points. Such an extension will be called a *free extension.*

From the method of constructing free extensions we have immediately the following theorem:

THEOREM 4.1. *The free extension of a partial plane to a complete plane is unique.*

For two free extensions of a partial plane to a complete plane can differ only in the order in which new elements are added. Every new element is constructed from points and lines of the given partial plane in a single way.

THEOREM 4.2. *If π_2 is a free extension of π_1 and π_3 is a free extension of π_2, then π_3 is a free extension of π_1. The free extension of π_2 to a complete plane is the free extension of π_1 to a complete plane.*

THEOREM 4.3. *If $\pi_1 \supset \pi_2$ and π_2 is complete in π_1 then the free extension π_1' of π_1 contains as a subplane the free extension π_2' of π_2. If π_2 is a proper part of π_1, then π_2 is a proper part of π_1'.*

Since π_2 is complete in π_1 its extension will include only new elements.

We may call the free extension of a partial plane π_1 to a complete plane simply the free extension of π_1. The free extension of $n-2$ points on a line and 2 points not on it will be written π^n.

THEOREM 4.4. *The free extension of any non-degenerate partial plane π_1 which is not a complete plane contains a π^4 as a subplane.*

Proof. Since π_1 is not a complete plane and is nondegenerate, it contains either (a) a pair of points P_1, P_2 not joined by any line or (b) a pair of lines L_1, L_2 which do not intersect. In case (b) suppose first that not every point of π_1 is on L_1 or L_2. Let such a point be P_1 and adjoin as P_2 the intersection of L_1 and L_2. Here P_1 and P_2 are two points of a free extension of π_1 not joined by a line. Second if every point of π_1 is on L_1 and L_2 and there are no further lines there must be at least two points each on L_1 and L_2 since π_1 is nondegenerate. If these are not all joined in pairs by lines take P_1 and P_2 as a pair not joined. If they are joined in pairs adjoin a point P_1 as intersection of two of these lines not intersecting on L_1 and L_2. Third suppose here P_1 and P_2 are the points of a free extension π_1 and case (b) is reduced to case (a).

Now take P_1 and P_2, two points of π_1 (or a free extension of π_1) not joined by a line. Suppose that π_1 contains another pair of points P_3 and P_4 which are not joined. Then extending π_1, if necessary, we have:

$$(4.1) \qquad L_2\colon P_1P_3B_2, \quad L_3\colon P_1P_4B_3, \quad L_4\colon P_2P_3B_3, \quad L_5\colon P_2P_4B_2, \quad M_3\colon B_2B_3.$$

This partial plane is necessarily complete in π_1 (or π_1 extended) since neither

P_1 and P_2 nor P_3 and P_4 are joined. Hence by Theorem 4.3 the free extension of π_1 contains the free extension of (4.1). But (4.1) is a free extension of the four points P_1, P_2, P_3, P_4 and hence by Theorem 4.2 the free extension of π_1 contains a π^4.

In any case π_1, being nondegenerate, or a simple extension not involving joining P_1 and P_2, must contain two points A_1 and A_2 which together with P_1 and P_2 form a quadrilateral. We may now suppose without loss of generality that π_1 is generated by P_1, P_2, A_1, A_2 and contains

$$(4.2)\qquad \begin{array}{lll} H_1\colon\ P_1A_1B_2, & H_3\colon\ P_2A_1B_3, & H_5\colon\ A_1A_2, \\ H_2\colon\ P_1A_2B_3, & H_4\colon\ P_2A_2B_2, & H_6\colon\ B_2B_3. \end{array}$$

Here the points and lines are distinct but (4.2) need not be complete in π_1 since H_5 and H_6 may intersect in π_1. Now adjoin $H_0\colon P_1P_2$ and the intersections of H_0 with H_5 and H_6 to π_1, giving $H_0\colon P_1P_2C_1C_2$ $H_2\colon A_1A_2C_1$ $H_6\colon B_2B_3C_2$. Now C_1 does not join B_2 nor does C_2 join A_1. Hence C_1, B_2, and C_2, A_1 may play the role of P_1, P_2 and P_3, P_4 in (4.1) and so the free extension of π_1 contains a π^4.

THEOREM 4.5. *A π^n, $n \geqq 4$, contains a π^{n+1}.*

Proof. Let π^n be generated by

$$(4.3)\qquad L_1\colon\ A_1A_2, \qquad L_6\colon\ A_3A_4, \cdots, A_n.$$

A partial free extension of (4.3) is

$$(4.4)\qquad \begin{array}{lll} L_1\colon\ A_1A_2B_1C_5, & L_4\colon\ A_2A_3B_3C_2, & M_1\colon\ B_1B_2C_1C_2, \\ L_2\colon\ A_1A_3B_2C_3, & L_5\colon\ A_2A_4B_2C_4, & M_2\colon\ B_1B_3C_3C_4, \\ L_3\colon\ A_1A_4B_3C_1, & L_6\colon\ A_3A_4B_1C_6A_5, \cdots, A_n, & M_3\colon\ B_2B_3C_5C_6, \\ & N_1\colon\ C_1C_3. & \end{array}$$

This contains the partial plane complete in (4.4),

$$(4.5)\qquad L_2\colon A_1A_3C_3,\ L_3\colon A_1A_4C_1,\ L_6\colon A_3A_4C_6A_5, \cdots, A_n,\ N_1\colon C_1C_3,$$

which is a free extension of

$$(4.6)\qquad N_1\colon\ C_1C_3, \qquad L_6\colon\ A_3A_4C_6A_5, \cdots, A_n,$$

and so π^n contains the free extension of (4.5) and hence of (4.6) which is a π^{n+1}.

Now in (4.4) the triangles $A_2A_3A_4$ and $B_1B_2B_3$ are perspective with respect to A_1 and the intersections of corresponding sides are C_2, C_4, and C_6. But in the free extension of (4.4) these three points are not collinear. Hence a π^n, $n \geqq 4$, is never Desarguesian and in consequence of Theorem 4.4 we have the following theorem.

THEOREM 4.6. *A free extension of a partial plane is never Desarguesian, provided the partial plane is nondegenerate and not a complete plane.*

If from a partial plane π_2 we remove a point which is on exactly two lines of π_2 or a line which contains exactly two points we obtain a partial plane π_1 of which π_2 is a free extension. This removal, or any succession of such removals, will be called a *free contraction*. If one partial plane may be obtained from another by any succession of free extensions or free contractions we say that the two partial planes are *free-equivalent*. It is easily shown that this is a true equivalence, being reflexive, symmetric, and transitive.

THEOREM 4.7. *A free contraction followed by a free extension may be replaced by a free extension followed by a free contraction.*

Proof. It is sufficient to prove this when there is only one step in the contraction and one in the extension. If we remove a point P and adjoin a line L we may first adjoin L and then remove P, since P was not used to define L, and the two lines which intersected in P have not been altered, nor does L pass through P. If we remove a point P and adjoin a point $Q \neq P$, we may first adjoin Q and then remove P, since the two lines for which Q is an intersection do not intersect before the removal of P. If we remove P and then adjoin $Q = P$, we may simply omit both steps or adjoin any element and then remove it. The situation is similar if the element first removed is a line.

A finite partial plane which has three or more points on every line and three or more lines through every point is said to be *confined*. Free contraction is not possible in a confined partial plane, or confined configuration as we may call it.

THEOREM 4.8. *A free plane does not contain any confined configuration.*

Proof. The generating partial plane (4.3) of a free plane π^n does not contain any confined configuration since there are only two lines. Hence a confined configuration α must contain elements of later stages. Consider an element of the last stage occurring in α. Suppose it to be a line L of the ith stage. The points on it must be of earlier stages. But in a free plane an ith stage line contains only two points of earlier stages, namely those defining it, and we have a contradiction. Similarly if an element of the last stage in α is a point we have a contradiction. Hence π contains no confined configuration.

We shall call a finite partial plane which contains no confined configuration an open configuration. Here is an example of an open configuration in which free contraction is not possible, but which becomes quite simple under free equivalence.

$$
\begin{array}{llll}
 & L_1\colon A_1A_2C_1, & L_3\colon A_1A_4B_3, & L_5\colon A_2A_4B_2, \\
(4.7) & L_2\colon A_1A_3B_2, & L_4\colon A_2A_3B_3, & L_6\colon A_3A_4C_2, \\
 & & L_7\colon B_2B_3C_3. &
\end{array}
$$

By free extension

$$
\begin{array}{lll}
L_1: A_1A_2C_1, & L_4: A_2A_3B_3D_1D_2D_3, & L_7: B_2B_3C_3, \\
L_2: A_1A_3B_2, & L_5: A_2A_4B_2, & M_1: A_4C_1D_1, \\
L_3: A_1A_4B_3, & L_6: A_3A_4C_2, & M_2: A_1C_2D_2, \\
 & M_3: A_1C_3D_3. &
\end{array} \tag{4.8}
$$

By free contraction we may remove first C_1, C_2, C_3 then M_1, M_2, M_3, L_1, L_6, L_7 and then B_2, B_3 and L_2 and L_5 leaving

$$
L_3: A_1A_4, \qquad L_4: A_2A_3D_1D_2D_3, \tag{4.9}
$$

the open configuration generating π^7. The general problem of free equivalence is treated in the following theorem.

THEOREM 4.9. *A nondegenerate open configuration is free equivalent to a configuration* (4.3) *which generates a* π^n.

Proof. Given an open configuration which is nondegenerate. Extending, if necessary, we may find a quadrilateral A_1, A_2, A_3, A_4 and hence by free equivalence we will have a configuration:

$$
\begin{array}{ll}
L_1: & A_1A_2(C_1)\cdots \qquad [B_1, \cdots, B_s], \\
L_2: & A_3A_4(C_1)\cdots, \\
L_3: & , \cdots, L_r: \ ,
\end{array} \tag{4.10}
$$

where the intersection C_1 of L_1 and L_2 may or may not occur, and the points $B_1, \cdots, B_s$ are on no line. We shall reduce the configuration in various ways by free equivalence, by reducing the number of lines, or keeping the number of lines fixed, always retaining two lines such as L_1 and L_2 containing a quadrilateral. This reduction process must come to an end in a finite number of steps and we shall show that the end is not reached as long as there are three or more lines, and then treat the cases with two lines.

First if there is a line L with no points, intersect L with L_1 and L_2 and then remove L. If there is a line L with exactly one point not equal to A_1, A_2, C_1, intersect L with L_1 and then remove L. If there is a line L with exactly one point, either A_1 or A_2, intersect L with L_2 and then remove L. If there is a line L with exactly the one point C_1 adjoin $M: A_1A_3$ if necessary, intersect with L and then remove L. Naturally we remove any line $L_3, \cdots, L_r$ with exactly two points. Hence we may suppose that either a reduction is possible or that each of the lines $L_3, \cdots, L_r$ contains at least three points. Now if some line, say L_3, contains a point $A_5 \neq A_1, A_2, A_3, A_4$ which occurs exactly once, then L_3 cannot contain both A_1 and A_2, and say it does not contain A_1. Then $M: A_1A_5$ will be a new line. By free extension (4.10) becomes

$$
\begin{aligned}
&L_1\colon\ A_1A_2(C_1)\cdots \qquad [B_1,\cdots,B_s],\\
&L_2\colon\ A_3A_4(C_1)\cdots D_1,\\
&L_3\colon\ A_5A_6\cdots,\ \cdots,L_r\colon\ ,\quad M\colon\ A_1A_5D_1.
\end{aligned}
\tag{4.11}
$$

By free contraction it becomes

$$
\begin{aligned}
&L_1\colon\ A_1A_2(C_1)\cdots \qquad [B_1,\cdots,B_s],\\
&L_2\colon\ A_3A_4(C_1)\cdots D_1,\\
&L_3\colon\ A_6\cdots,\ \cdots,L_r\colon\ .
\end{aligned}
\tag{4.12}
$$

Hence a point A_5 occurring exactly once on L_3 may be replaced by a point occurring exactly once on L_2 (or similarly on L_1). Since this process is clearly reversible, we may shift the points of single occurrence at will and we shall suppose that they are all on one line, either L_1 or L_2 according to preference. Note that this does not involve any increase in the number of points. We remove any point occurring exactly twice, except one of A_1, A_2, A_3, A_4. If one of these occurs exactly twice, suppose this to be A_1 occurring a second time on L_3 and that the points of single occurrence are on L_2. Now L_3 cannot contain C_1 and if it contains neither A_3 nor A_4, then $L_3\colon A_1A_5A_6\cdots$ where A_3, A_4, A_5, A_6 form a quadrilateral and we may reduce by dropping A_1 and keeping L_2, L_3 in the role of L_1, L_2. If L_3 contains one of A_3, A_4 say A_3, then $L_3\colon A_1A_3A_5\cdots$ and this reduction fails. Here if C_1 occurs, then in its third occurrence $L_4\colon C_1A_6A_7\cdots$ and we may take A_3, A_4, A_6, A_7 as a quadrilateral on L_2 and L_4 and reduce by dropping A_1. If C_1 does not occur, then A_5 occurs on at least two lines besides L_3 and at least one of these is not through A_4, say $L_5\colon A_5A_6\cdots$. Here we may take A_3, A_4, A_5, A_6 as a quadrilateral and drop A_1 for a reduction. Hence if there are more than two lines and (4.10) is not reducible by one of these devices, we may suppose that every point occurs either once or at least three times, that the points of single occurrence are on L_1 or L_2, and that each of $L_3,\cdots,L_r$ contains at least three points. Since there are at least three points on L_3 and since each of these occurs at least three times, there must be a fourth line to yield enough incidences. Now since L_3 and L_4 have three or more points each, points occurring three or more times must contain a quadrilateral of points each occurring three or more times. Hence shifting from A_1, A_2, A_3, A_4 to this quadrilateral, if necessary, we may assume that A_1, A_2, A_3, A_4 each occurs at least three times in (4.10) or that further reduction is possible. Now if in (4.10) we suppress all points of single occurrence, the remaining configuration must, by hypothesis, be open. But this leaves $L_3,\cdots,L_r$ with at least three points each, L_1 with at least A_1, A_2 and L_2 with at least A_3, A_4 and every point occurring three or more times. Hence one of L_1, L_2, say L_1, can have no more than the two points A_1, A_2. Now in (4.10), without suppression of points of single occurrence, the single points may be supposed shifted to L_2. Since there are no

points of double occurrence, this leaves L_1 with exactly the two points A_1, A_2. L_3 and L_4 will certainly contain a quadrilateral and so we may reduce by dropping L_1 and still keep two lines with a quadrilateral. This completes reduction of (4.10) by free equivalence to a configuration with two lines containing a quadrilateral,

$$(4.13) \qquad L_1: \ A_1A_2D_1 \cdots D_t \ [B_1, \cdots, B_s], \qquad L_2: \ A_3A_4 \cdots A_u,$$

where C_1 has been removed if it occurred. By free extension

$$(4.14) \qquad \begin{aligned} &L_1: \ A_1A_2D_1 \cdots D_t \ [B_2, \cdots, B_s], \qquad L_3: \ A_1B_1A_{u+1}, \\ &L_2: \ A_3A_4 \cdots A_uA_{u+1}A_{u+2}, \qquad L_4: \ A_2B_1A_{u+2}. \end{aligned}$$

By free contraction we remove B_1 and then L_3, L_4, yielding

$$(4.15) \qquad L_1: \ A_1A_2D_1 \cdots D_t \ [B_2, \cdots, B_s], \qquad L_2: \ A_3A_4 \cdots A_{u+2}.$$

Thus each B may be replaced by two A's on L_2 until we have

$$(4.16) \qquad L_1: \ A_1A_2D_1 \cdots D_t, \qquad L_2: \ A_3A_4 \cdots A_v.$$

Now we may use a free extension

$$(4.17) \qquad \begin{aligned} &L_1: \ A_1A_2D_1 \cdots D_t, \qquad L_3: \ A_1A_3E_1, \\ &L_2: \ A_3A_4 \cdots A_vA_{v+1}, \qquad L_4: \ A_2A_4E_1, \\ &\qquad\qquad L_5: \ D_tE_1A_{v+1}, \end{aligned}$$

where by free contraction we may remove D_t, then L_5, then E_1, and finally L_3 and L_4 leaving

$$(4.18) \qquad L_1: \ A_1A_2D_1 \cdots D_{t-1}, \qquad L_2: \ A_3A_4 \cdots A_{v+1}.$$

In this way we may replace every D by another A on L_2 reaching

$$(4.19) \qquad L_1: \ A_1A_2, \qquad L_2: \ A_3A_4 \cdots A_n,$$

which we may call the free n-point. This completes the proof of the theorem.

In any configuration let P be the number of points, L the number of lines, and I the number of incidences.

THEOREM 4.10. *The number $2(P+L)-I$ is invariant under free equivalence.*

Proof. In free extension, adding a point we increase P by 1 and I by 2. Adding a line, we increase L by 1 and I by 2. In free contraction we decrease by similar amounts. In all cases $2(P+L)-I$ remains unchanged.

THEOREM 4.11. *An open configuration α which is nondegenerate is free equivalent to the free n-point if and only if*

$$2(P + L) - I = n + 4.$$

Proof. By Theorem 4.9, α is free-equivalent to some free n-point and by Theorem 4.10, $2(P+L)-I$ is invariant under free equivalence. Now for (4.19) $P=n$, $L=2$, $I=n$ and $2(P+L)-I=n+4$.

THEOREM 4.12. *A free n-point and a free m-point are not free equivalent if $m \neq n$. The planes π^m and π^n are not isomorphic.*

An immediate consequence of Theorems 4.10 and 4.11 is that the free m-point and the free n-point are not free equivalent if $m \neq n$. If π^m and π^n were isomorphic we could consider both as the same plane generated in one way by a free m-point and in another by a free n-point. Let C_1 be a configuration generated by the free m-point large enough to include the generating free n-point. Then extend C_1 by generating from the free n-point so that the generating free m-point is constructible from the free n-point. Call the extended configuration C_2. Now from its method of construction C_2 may be considered as generated either by the free m-point or by the free n-point. By assumption the whole plane is the free extension of both the m-point and the n-point. Hence by free contraction C_2 is free-equivalent to both the m-point and the n-point, and they are free-equivalent to each other. But this is a contradiction and hence π^m and π^n cannot be isomorphic.

THEOREM 4.13. *A plane containing no confined configuration and generated by a finite configuration is a free plane π^n.*

COROLLARY. *The finitely generated subplanes of a free plane are free.*

Proof. Let π contain no confined configurations, and be generated by a finite configuration α. In the extension of α the free adjunction of a point or line leaves $2(P+L)-I$ invariant, while the adjunction of a point on three or more lines or a line through three or more points decreases the value of $2(P+L)-I$. Now a configuration for which $2(P+L)-I \leqq 7$ cannot by Theorem 4.11 be an open configuration. Hence in any (finite) extension of α in π, $2(P+L)-I \geqq 8$ and consequently in generating π from α there are only a finite number of adjunctions which are not free. Hence π may be considered the free extension of some finite extension β of α. But β is an open configuration and hence a finite extension γ of β is a free extension of a free n-point and hence π, as a free extension of γ, is a free plane π^n.

It is conceivable that Theorem 4.13 would not be true without the assumption of the finiteness of α. It is possible that the generators of a plane might be so scattered and so interrelated that it might be impossible to find points on a single line and two points off it which would generate the entire plane without any relation between them.

It is easy to prove by the method of Theorem 4.8 that a free extension of a partial plane contains no confined configuration except those included entirely in the partial plane. Thus any collineation or duality of the free

extension of a confined configuration maps the confined configuration into itself. In this way it is not too difficult to construct planes which are not self-dual, planes with only the identity as a collineation, and so on. Further investigation along these lines might be rewarding.

5. **Introduction of coordinates.** Classical investigations(6) have shown that in a Desarguesian plane we may introduce homogeneous coordinates from a skew-field and thereafter proceed by the methods of analytic geometry. For a given skew-field F, we denote points by $P(\lambda x_1, \lambda x_2, \lambda x_3)$, x_1, x_2, x_3 not all zero, $\lambda \neq 0$ arbitrary. We denote lines by $L[a_1\sigma, a_2\sigma, a_3\sigma]$, a_1, a_2, a_3 not all zero, and $\sigma \neq 0$ arbitrary. Here $P \subset L$ if and only if

$$x_1a_1 + x_2a_2 + x_3a_3 = 0. \tag{5.1}$$

Given any skew-field F it is not difficult to verify that points and lines so defined form a Desarguesian plane. We shall distinguish between an analytic plane and a Desarguesian plane until the equivalence has been shown here. It is not the purpose of this paper to duplicate this classic result, but it will be incidental to other investigations.

In an analytic plane it is easy to verify that the projective transformations

$$\begin{aligned} \rho x_1 &= b_{11}x_1' + b_{12}x_2' + b_{13}x_3', \\ \rho x_2 &= b_{21}x_1' + b_{22}x_2' + b_{23}x_3', \\ \rho x_3 &= b_{31}x_1' + b_{32}x_2' + b_{33}x_3', \end{aligned} \tag{5.2}$$

with $\rho \neq 0$ and the matrix (b_{ij}) of rank three, form a group of collineations and that by an appropriate collineation the vertices of any quadrilateral may be taken into $A(0, 1, 0)$, $B(1, 0, 0)$, $C(0, 0, 1)$, and $D(1, 1, 1)$. In the following section we shall select an arbitrary quadrilateral in any projective plane π and introduce coordinates with respect to which it plays a role similar to that of $ABCD$ in the analytic plane.

Let us select any two distinct points A and B in the projective plane π. If from π we remove the line L_∞ joining A and B and the points on L_∞, the remaining plane π^* is a Euclidean plane. We shall use A and B as centers of perspectivities to introduce coordinates for the points of π^*. No ambiguity will arise if, as we shall find convenient on occasion, we speak of π and π^* as the same plane so long as L_∞ is fixed.

Let the lines of the pencil through A in π^* be denoted by $x = x_0, \cdots, x = x_{n-1}$ where n is their cardinal number and $x_0, \cdots, x_{n-1}$ are n different symbols. Similarly denote the lines of the pencil through B by $y = y_0, \cdots, y = y_{n-1}$. (The cardinal number of the y's is necessarily the same as that of the x's.) Then through every point P of π^* there is exactly one line $x = x_i$ and one line $y = y_j$. Denote P by (x_i, y_j). As in Figure 3 we see that on an arbitrary line L of π^* not through A or B there are points (x_i, y_i) where each x and each y

(6) D. Hilbert [1], K. G. C. von Staudt [1], vol. 2, pp. 166 et seq.

occurs exactly once. Henceforward suppose that the symbols $x_0, \cdots, x_{n-1}$ are the same as $y_0, \cdots, y_{n-1}$, though not necessarily in the same order. Then an arbitrary line L not through A or B is associated with the permutation $\begin{pmatrix} \cdots, x_i, \cdots \\ \cdots, y_i, \cdots \end{pmatrix}$ where the (x_i, y_i) are the points of L. This expresses the fact (see Figure 3) that L determines a 1-1 correspondence between the lines of the pencil through A and the lines of the pencil through B. Without confusion we may write

$$L = \begin{pmatrix} \cdots, x_i, \cdots \\ \cdots, y_i, \cdots \end{pmatrix} \tag{5.3}$$

since distinct lines contain at most one point in common and hence are associated with distinct permutations.

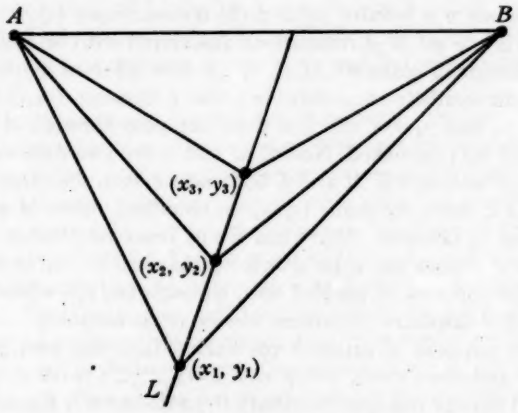


FIG. 3

THEOREM 5.1. *The projection* $L \to_A M \to_B L$ *induces the permutation* $(x_i, y_i) \to (x_j, y_j)$ *of the points of* L *where* $\begin{pmatrix} \cdots, x_i, \cdots \\ \cdots, x_j, \cdots \end{pmatrix} = ML^{-1}$.

Proof. The perspectivity $L \to_A M$ takes (x_i, y_i) of L into (x_i, z_i) of M and the perspectivity $M \to_B L$ takes (x_i, z_i) of M into (x_j, y_j) of L where $y_j = z_i$. But $M = \begin{pmatrix} \cdots, x_i, \cdots \\ \cdots, z_i, \cdots \end{pmatrix}$ and $L^{-1} = \begin{pmatrix} \cdots, y_j, \cdots \\ \cdots, x_j, \cdots \end{pmatrix}$ where $z_i = y_j$ and hence $ML^{-1} = \begin{pmatrix} \cdots, x_i, \cdots \\ \cdots, x_j, \cdots \end{pmatrix}$.

We have shown that in an arbitrary Euclidean plane π^* the points may be represented by (x_i, y_j) and the lines by $x = c_i$, $y = c_j$ and permutations L of n symbols $c_0, \cdots, c_{n-1}$. What are the properties that characterize the set S of permutations $[L]$? This question is answered by the following theorem.

THEOREM 5.2. *A set* S *of permutations on* n *letters* $c_0, \cdots, c_{n-1}$ *will be associated with the lines of a Euclidean plane* π^* *not through its centers of perspectivity if and only if* (a), (b), *and* (c) *or* (c′) *hold.*

(a) *S is doubly transitive on $c_0, \cdots, c_{n-1}$.*

(b) *If M and L are different permutations of S, then ML^{-1} fixes at most one letter.*

(c) *If L does not take c_i into c_j there is one and only one permutation M which does take c_i into c_j such that ML^{-1} displaces all letters.*

(c′) *n is finite.*

REMARK. Note that an interchange of the centers of perspectivity, A and B, replaces S by S^{-1}, the set of inverses of elements in S, and that S^{-1} satisfies conditions (a), (b), and (c) if and only if S satisfies them.

Proof. In this proof we shall use conditions (a), (b), and (c). See appendix I for a proof that (a), (b), and (c′) imply (c) and also for counter-examples showing that when n is infinite (a) and (b) do not imply (c).

Suppose S is the set of permutations associated with lines L not through A or B in a Euclidean plane π^*. If x_1, x_2 are two different symbols and y_1, y_2 are two different symbols then there is a line L through (x_1, y_1) and (x_2, y_2), and since $x_1 \neq x_2$ and $y_1 \neq y_2$ this line does not pass through A or B. Hence $L = \left(\begin{smallmatrix} x_1, x_2, \cdots \\ y_1, y_2, \cdots \end{smallmatrix}\right)$ and (a) is satisfied. Now if M and L are two different lines ML^{-1} fixes a letter c_i if and only if M and L both take c_i into the same letter, say c_j, whence M and L have the point (c_i, c_j) in common. Since M and L have at most one point in common ML^{-1} can fix at most one letter. Hence (b) is satisfied. Now if L does not take c_i into c_j, that is if (c_i, c_j) is not on L, then there is exactly one line M parallel to L through (c_i, c_j), whence M takes c_i into c_j and ML^{-1} displaces all letters. Hence (c) is satisfied.

Conversely suppose S satisfies conditions (a), (b), and (c). Construct *points* (x_i, y_j) and *lines* $x = c_i$, $y = c_j$ and $L = \left(\begin{smallmatrix} \cdots, x_i, \cdots \\ \cdots, y_i, \cdots \end{smallmatrix}\right)$ with L from S. Then the points and lines of this system satisfy the axioms for a Euclidean plane π^*. Using (a) we may verify that there is a line through two arbitrary points. From (b) we may show that two lines intersect in at most one point. From (c) it follows that through a given external point there is exactly one line parallel to a given line. In the verification of these axioms, the lines $x = c_i$ and $y = c_j$ form two sets of parallels and assure us that there is a line through (x_1, y_1) and (x_2, y_2) even if $x_1 = x_2$ or $y_1 = y_2$.

In a plane any two lines contain the same number of points (here n for the Euclidean plane, $n+1$ for the projective plane). Hence if two planes are isomorphic the number of points per line must be the same for both. Since in Theorem 5.2 the symbols $c_0, \cdots, c_{n-1}$ are any n distinct symbols, without loss of generality we may assume that the same symbols are used in representing any two planes which have the same number of points on a line.

THEOREM 5.3. *The plane π_1 represented on the pencils through A_1 and B_1, using symbols $c_0, \cdots, c_{n-1}$, will be isomorphic to the plane π_2 represented on the pencils through A_2 and B_2, using symbols $c_0, \cdots, c_{n-1}$, with $A_1 \rightleftarrows A_2$ and $B_1 \rightleftarrows B_2$, if and only if there exist two permutations X and Y of $c_0, \cdots, c_{n-1}$ such that*

$X^{-1}S_1Y=S_2$, where S_1 is the set of L permutations of π_1 and S_2 is the set of L permutations of π_2. Here $X=\begin{pmatrix}\cdots,c_i,\cdots\\ \cdots,c_j,\cdots\end{pmatrix}$ where the line $x=c_i$ of the A_1 pencil is mapped onto the line $x=c_j$ of the A_2 pencil, and $Y=\begin{pmatrix}\cdots,c_r,\cdots\\ \cdots,c_s,\cdots\end{pmatrix}$ where the line $y=c_r$ of the B_1 pencil is mapped onto the line $y=c_s$ of the B_2 pencil.

Proof. Suppose there is an isomorphism $\pi_1\rightleftarrows\pi_2$, $A_1\rightleftarrows A_2$, $B_1\rightleftarrows B_2$. Then the pencil through A_1 is mapped onto the pencil through A_2, and that through B_1 mapped onto that through B_2. Hence correspondences $x=c_i\rightleftarrows x=c_j$ and $y=c_r\rightleftarrows y=c_s$ are determined. Write $X=\begin{pmatrix}\cdots,c_i,\cdots\\ \cdots,c_j,\cdots\end{pmatrix}$, $Y=\begin{pmatrix}\cdots,c_r,\cdots\\ \cdots,c_s,\cdots\end{pmatrix}$. Then $P_1(c_i, c_r)\rightleftarrows P_2(c_j, c_s)$. Here if $P_1\subset L_1$ and $L_1\rightleftarrows L_2$, then $P_2\subset L_2$. Hence if the permutation L_1 takes c_i into c_r then $X^{-1}L_1Y$ takes c_j into c_s whence $L_2=X^{-1}L_1Y$ as was to be shown. Conversely suppose that $X^{-1}S_1Y=S_2$. Then let us set up the correspondences $A_1\rightleftarrows A_2$, $B_1\rightleftarrows B_2$, $x=c_i\rightleftarrows x=c_j$, $y=c_r\rightleftarrows y=c_s$ and $P_1(c_i, c_r)\rightleftarrows P_2(c_j, c_s)$. Also put $L_1\rightleftarrows L_2$ if $L_2=X^{-1}L_1Y$. Here $P_1\subset L_1$ if and only if $P_2\subset L_2$. Hence this is an isomorphism between π_1 and π_2, which maps A_1 onto A_2 and B_1 onto B_2.

We may use the permutations S associated with a plane π to introduce as another representation of π a ternary ring. This ternary ring, as will be proved later, reduces to the usual coordinate ring in an analytic plane, but, without conditions of Desarguesian type, it will have very few of the ordinary ring properties of a skew-field. Among the various equivalent representations of π which Theorems 5.2 and 5.3 yield we may, by taking $X=L$, $Y=I$, obtain a representation in which an arbitrary line $L\rightleftarrows L^{-1}LI=I$ is represented by the identity. Now take two arbitrary symbols as $c_0=0$, and $c_1=1$. We keep unchanged the representation of the lines $x=c_i$, $y=c_j$. The remaining lines M we use to define a ternary operation $y=x\cdot m\circ b$ on the symbols $0, 1, c_2, \cdots, c_{n-1}$. To a given line M correspond parameters m, b in the following way:

(1) If $M=\begin{pmatrix}0,1,\cdots\\ 0,s,\cdots\end{pmatrix}$, then for M take $m=s$, $b=0$.

(2) If M goes through $(0, a)$, $a\neq 0$, and is parallel to $\begin{pmatrix}0,1,\cdots\\ 0,s,\cdots\end{pmatrix}$ then for M take $m=s$, $b=a$. If (x, y) is any point on M define the ternary operation as $y=x\cdot m\circ b$. Here a point is represented by a pair of coordinates (x, y) and a line contains all those points bound by the ternary relation $y=x\cdot m\circ b$. This operation is defined for all x, m, b selected from the system $0, 1, c_2, \cdots, c_{n-1}$ except for triples with $m=0$. We may extend slightly by defining $x\cdot 0\circ b=b$ for any x and thus include the lines $y=c_j$ in the general form. In the analytic planes $(0, b)$ is the intercept of M on the y-axis, m is the slope of M, and the equation of M is $y=xm+b$. What in general are the properties of this ternary operation and when does such a ternary operation define a plane?

THEOREM 5.4(7). *A ternary operation $a\cdot b\circ c$ uniquely defined for all ordered triples a, b, c containing elements 0 and 1 may be derived from the incidences of a*

(7) Similar results, as yet unpublished, have been obtained by D. T. Perkins.

Euclidean plane having the following properties:

(1) $0 \cdot m \circ c = a \cdot 0 \circ c = c$, $1 \cdot m \circ 0 = m$, $a \cdot 1 \circ 0 = a$.

(2) $x \cdot m \circ b = c$, *for given* $m \neq 0$, b, c, *has a unique solution* x.

(3) $a \cdot m \circ z = c$, *for given* a, m, c, *has a unique solution* z.

(4) $a \cdot z \circ b = c$, *for given* $a \neq 0$, b, c, *has a unique solution* z.

(5) $x \cdot m_1 \circ b_1 = x \cdot m_2 \circ b_2$ *has a unique solution* x *if* $m_1 \neq m_2$.

(6) *The pair of equations* $x_1 \cdot m \circ b = y_1$, $x_2 \cdot m \circ b = y_2$ *has a unique solution* $m = s$, $b = a$ *given* $x_1 \neq x_2$, $y_1 \neq y_2$.

For each choice of a fundamental quadrilateral A, B, $O = (0, 0)$, $I = (1, 1)$ *in* π *the ternary operation is uniquely determined. Conversely a ternary operation on a system* T *with the above properties defines a Euclidean plane with points* (a, b), $a, b \subset T$, *and lines* $x = c$ *and* $y = x \cdot m \circ b$.

Proof. As was shown above, a plane may be represented as in Theorem 5.2, taking one line L as the identity, the coordinates being from a system $T(0, 1, c_2, \cdots, c_{r-1})$. The points are (c_i, c_j) and the lines are $x = c_i$, $y = x \cdot m \circ b$. Here $a \cdot 0 \circ c = c$ is part of the definition of the ternary operation and $0 \cdot m \circ c = c$ states that $(0, c)$ is on $y = x \cdot m \circ c$. $1 \cdot m \circ 0 = m$ states that $(1, m)$ is on $\binom{0, 1, \cdots}{0, m, \cdots}$ if $m \neq 0$ while $1 \cdot 0 \circ 0 = 0$ is a special case of $a \cdot 0 \circ c = c$. $a \cdot 1 \circ 0 = a$ states that $\binom{0, 1, \cdots}{0, 1, \cdots}$ is the identical permutation. So much for property (1). (2) states that $y = c$ intersects $\binom{0, \cdots}{b, \cdots}$ in exactly one point. In (3) if $m = 0$, $a \cdot m \circ z = z = c$ and z is unique. If $m \neq 0$ it states that there is a unique parallel to $\binom{0, 1, \cdots}{0, m, \cdots}$ through (a, c). (4) states that if $a \neq 0$ and $b = c$ then $z = 0$ is a solution, whereas if $a \cdot m \circ b = c$ with $m \neq 0$ we have $\binom{0, a, \cdots}{b, c, \cdots}$ parallel to $\binom{0, 1, \cdots}{0, m, \cdots}$, hence $c \neq b$. If $b \neq c$ there is a unique parallel $\binom{0, 1, \cdots}{0, m, \cdots}$ through $(0, 0)$ to $\binom{0, a, \cdots}{b, c, \cdots}$, and $z = m$ is the unique solution. (5) states that the lines $y = x \cdot m_1 \circ b_1$ and $y = x \cdot m_2 \circ b_2$, if $m_1 \neq m_2$, are not parallel and intersect in a unique point (x, y). (6) states that there is a unique line through (x_1, y_1) and (x_2, y_2) if $x_1 \neq x_2$, $y_1 \neq y_2$. Hence the ternary operation defined by a plane has properties (1) to (6) inclusive.

Conversely suppose that in a system $T(0, 1, c_2, \cdots, c_{n-1})$ a ternary operation $a \cdot b \circ c$ is defined satisfying properties (1) to (6). Let us define *points* (c_i, c_j) and *lines* $x = c_i$, $y = c_j = x \cdot 0 \circ c_j$, and lines $M(m, b) = \binom{\cdots, c_i, \cdots}{\cdots, c_j, \cdots}$ if $c_j = c_i \cdot m \circ b$ for every $m \neq 0$, and every b. Since $x \cdot m \circ b$ is uniquely defined for every x in M, c_i takes on every value exactly once. From property (2) every c_j occurs exactly once in M and so M is a permutation of the c's. From property (6) the permutations M are doubly transitive. From property (5) $M_1 M_2^{-1}$ fixes exactly one letter if the slopes m_1 and m_2 are different. If $m_1 = m_2 = m$, $b_1 \neq b_2$, $x \cdot m \circ b_1 \neq x \cdot m \circ b_2$ since from property (3) $a \cdot m \circ z = c$ has a unique solution, and in this case $M_1 M_2^{-1}$ fixes no letter. If $m_1 = m_2$ and $b_1 = b_2$, the lines M_1 and M_2 are identical. From the definition of the ternary operation and property (2) it follows that no line $y = x \cdot m \circ b$ with $m \neq 0$ can be parallel to a line $x = c$ or $y = c$. Hence through an external point there is exactly one parallel to a

line of slope m whose slope is m (property 3), and no parallel with a different slope. Hence the conditions (a), (b), (c) of Theorem 5.2 are all satisfied and the ternary relation does in fact define a plane. In addition, one of the lines, $y=x\cdot 1\circ 0$, is the identity $y=x$. Note that in this coordinatization of the plane not only A and B, but also (0, 0) and (1, 1) play a special role and determine from the plane the ternary operation to within isomorphism. Different choices of the fundamental quadrilateral may lead to systems not isomorphic in any algebraic sense. For examples, see appendix II.

We may define both addition and multiplication as special cases of the ternary operation.

Addition: $a+b=c$ if and only if $a\cdot 1\circ b=c$.

Multiplication: $ab=c$ if and only if $a\cdot b\circ 0=c$. Here we have $0+a=a+0=a$ and the existence of unique u, v such that $u+a=b$, $a+v=b$, though not even the right and left negatives of a need be equal. Also $a0=0a=0$ and $a1=1a=a$, while for $a\neq 0$, $ua=b$, $av=b$ have unique solutions, though not even the right and left inverses need be the same. In general no other ring properties hold, nor will it be true that $x\cdot m\circ b=xm+b$.

In an affine analytic plane lines $y=xm+b$ with $m\neq 0$ may be associated with the transformations $x\to xm+b$ of the linear group. Here the permutations $\begin{pmatrix}\cdots,&x,&\cdots\\ \cdots,&y,&\cdots\end{pmatrix}$ will be the elements of the linear group. Hence when skew-field coordinates have been introduced, the permutations of Theorem 5.2 form a group. But it is possible for these permutations to form a group even in certain non-analytic planes. See appendix II for an example of such a plane.

THEOREM 5.5. *If the A_1-B_1 permutations of a plane π_1 form a group G, the A_2-B_2 permutations of a plane π_2 isomorphic to π_1 with $A_2\rightleftarrows A_1$, $B_2\rightleftarrows B_1$ form a coset of a transform of G in the symmetric group on $c_0, \cdots, c_{n-1}$.*

Proof. The permutations S_1 of π_1 are the group G. By Theorem 5.3 the permutations of π_2 are $S_2=X^{-1}S_1Y$ for appropriate permutations X and Y. Hence $S_2=X^{-1}GY=X^{-1}GX\cdot X^{-1}Y$ is a left coset of $X^{-1}GX$, a transform of G. Also $S_2=X^{-1}Y\cdot Y^{-1}GY$ is a right coset of $Y^{-1}GY$.

THEOREM 5.6([8]). *A permutation group G on letters $c_0, \cdots, c_{n-1}$ such that*

(α) *G is doubly transitive,*

(β) *only the identity fixes two letters,*

(γ) *at most one element taking c_i into c_j displaces all letters, or*

(γ') *n is finite*

will be the permutations S of a Euclidean plane π^*. The identity and the elements of G which displace all letters form a simply transitive normal abelian subgroup H.

([8]) When n is finite, this theorem rests on a theorem of Frobenius, A. Speiser [1, p. 200], depending on the theory of group characters, or on the Sylow theorems as proved in R. D. Carmichael [1, p. 145]. Compare Lemma 8 with page 25 in G. Thomsen [1].

Proof. In a group, clearly conditions (α) and (β) are equivalent to conditions (a) and (b) of Theorem 5.2. As is shown in appendix I, (a), (b) and (γ') imply (c). When n is infinite condition (γ) is only half of condition (c). It is the "only one" part of the condition. We shall see in Lemma 5 below that the part "there exists one" of condition (c) is a consequence of (α) and (β) for a group G. The second counter-example in appendix II shows that this is *not* a consequence of conditions (a) and (b). The author does not know whether (α) and (β) for a group G imply (γ). A counter-example, if it exists, would have no particular geometric interpretation and would be of relatively little interest here. When Lemma 5 has been reached, we shall have shown that the permutations of G *are the permutations* S of a plane.

LEMMA 1. *There exists one and only one element of order two in* G *which interchanges a pair of letters* (i, j).

Since G is doubly transitive there exists an element $g=\left(\begin{smallmatrix}\cdots & i & j & \cdots\\ \cdots & j & i & \cdots\end{smallmatrix}\right)$. If h also interchanges i and j then gh^{-1} fixes both i and j, whence $gh^{-1}=1$, $g=h$. Hence g is unique. Also g^2 fixes both i and j and hence $g^2=1$.

LEMMA 2. *The elements of order* 2 *in* G *are a conjugate class.*

An element of order 2 fixes at most one letter. If $n=2$ there is only one element of order 2 and the lemma holds. If $n\geqq 3$ then g and h of order 2 must both displace some letter i; $g=(ij)\cdots$, $h=(ik)\cdots$. G contains an element $x=\left(\begin{smallmatrix}i & j & \cdots\\ i & k & \cdots\end{smallmatrix}\right)$ and $x^{-1}gx=(ik)\cdots=h$. If an element fixes a letter, then any conjugate will also fix some letter. Hence we may subdivide into two cases: Case I, The elements of order 2 displace all letters; Case II, Every element of order 2 fixes a letter.

LEMMA 3. *In Case* II *there is one and only one element of order* 2 *fixing a given letter.*

In Case II, as before, $g=(ij)\cdots$ is transformed into $h=(ik)\cdots$ by $x=\left(\begin{smallmatrix}i & j & \cdots\\ i & k & \cdots\end{smallmatrix}\right)$. But here if g and h both fix the letter s, then x must also fix s. Hence $x\neq 1$ fixes both i and s which contradicts assumption (β). If g fixes a letter s and t is any other letter, an x taking s into t will transform g into an h of order 2 which fixes t.

LEMMA 4. *The product of two different elements of order* 2 *is an element displacing all letters.*

Suppose to the contrary that $g^2=1$, $h^2=1$, $g\neq h$, and gh fixes the letter i. If either g or h fixes i then the other must also, but this is impossible by Lemma 3. If both g and h displace i, then the letter j into which g takes i must be taken into i by h and $g=(ij)\cdots$, $h=(ji)$, whence by Lemma 1 $g=h$, contrary to assumption. Hence it is impossible that gh could fix any letter.

LEMMA 5. *In G there is at least one element, displacing all letters, which takes a given i into a given $j \neq i$.*

In Case I there is an element of order 2, displacing all letters, which takes i into j. In Case II there is an element g of order 2 fixing i, and an element h of order 2 taking i into j. Here gh takes i into j and by Lemma 4 displaces all letters.

As yet we have made no use of assumption (γ). From now on it is necessary to use it. From Lemma 5 and (γ) there is exactly one element in G, displacing all letters, taking i into j. Now if L is an element taking k into j, $k \neq i$, then let T be the element displacing all letters taking i into k. Then take $M = TL$. Here M takes i into j and $ML^{-1} = T$ displaces all letters. Since T is unique, M also is unique. This proves property (c). Hence the elements of G form the permutations S of a Euclidean plane π^*.

LEMMA 6. *In Case* I *every element displacing all letters is of order* 2 *and these together with the identity form a normal abelian subgroup.*

Since there is an element of order 2, displacing all letters, taking i into j, this must by (γ) be the only element, displacing all letters, taking i into j. But since i and j are arbitrary, every element displacing all letters must be of order 2. If $g^2 = 1$, $h^2 = 1$, then if $g = h$, $gh = 1$, while if $g \neq h$, gh displaces all letters and hence $(gh)^2 = 1$, whence $hg = gh$, and the elements of order 2 together with the identity form an abelian subgroup H. By Lemma 2, H is a normal subgroup. This completes the proof of the theorem for Case I.

LEMMA 7. *In Case* II *an element g displacing all letters may be written as the product $g = ab$ of two elements of order* 2 *where either a or b may be chosen arbitrarily.*

Given g displacing all letters and $a \neq 1$, $a^2 = 1$. Suppose that a fixes the letter i and that g takes i into j. Let $b = (ij) \cdots$. Then ab takes i into j and by Lemma 4 displaces all letters, whence by (γ), $g = ab$. Similarly, given g and b fixing k, if g takes i into k and $a = (ik) \cdots$, then $g = ab$.

LEMMA 8. *In Case* II *the product abc of three elements of order* 2 *is again an element of order* 2, *and $abc = cba$.*

Here the lemma is trivial if $b = a$. If $b \neq a$, then $ab = g = dc$ where, by Lemma 7, $d^2 = 1$. Hence $abc = dc^2 = d$. Since $d = d^{-1}$, $abc = c^{-1}b^{-1}a^{-1} = cba$.

LEMMA 9. *In Case* II *the elements displacing all letters, together with the identity, form a normal abelian simply transitive subgroup H.*

By Lemmas 4 and 7 these are the elements of the form ab where a and b are of order 2, and hence are a normal set. By Lemma 5 this set is transitive. Clearly the inverse of ab is ba. To show that these form an abelian subgroup

we must show $g_1g_2=1$ or g_3 and $g_1g_2=g_2g_1$. Here $g_1g_2=a_1b_1a_2b_2$ and by Lemma 7 $g_1=da_2$, whence $g_1g_2=db_2=1$ or g_3. By Lemma 8 $g_1g_2=(a_1b_1a_2)b_2=(a_2b_1a_1)b_2=a_2(b_1a_1b_2)=a_2(b_2a_1b_1)=g_2g_1$. This completes the proof of Theorem 5.6.

THEOREM 5.7. *Under the assumptions of Theorem* 5.6 *in* π^* *we have* $x\cdot m\circ b=xm+b$ *and addition and multiplication satisfy*:

A1. $(a+b)+c=a+(b+c)$,
A2. $b+a=a+b$,
A3. $a+0=0+a=a$,
A4. $a+(-a)=(-a)+a=0$,
M1. $(ab)c=a(bc)$,
M3. $a1=1a=a$,
M4. $aa^{-1}=a^{-1}a=1$ *if* $a\neq 0$,
M5. $a0=0a=0$,
D1. $(a+b)m=am+bm$,
E3. $xr=xs+t$ *has a unique solution* x *if* $r\neq s$.

Proof. Following Theorem 5.4 the identity is $I=\binom{0,1,\cdots}{0,1,\cdots}$. If $A_a=\binom{0,\cdots,u,\cdots}{a,\cdots,v,\cdots}$ is the permutation taking 0 into a and displacing all letters then $v=u\cdot 1\circ a=u+a$. If $M_m=\binom{0,1,\cdots,w,\cdots}{0,m,\cdots,z,\cdots}$ then $z=w\cdot m\circ 0=wm$. An arbitrary permutation may be written in the form $L=\binom{0,1,\cdots}{b,c,\cdots}=\binom{0,1,\cdots}{0,m,\cdots}\binom{0,\cdots,m,\cdots}{b,\cdots,c,\cdots}=M_m\cdot A_b$ whence $x\cdot m\circ b=xm+b$.

From Theorem 5.6-the A_a's and I form an abelian subgroup H whence A1, A2, A3, and A4 hold. The M_m's and I form the subgroup of G fixing the letter O, whence multiplication forms a group and M1, M3, M4 hold. To show that the one-sided distributive law D1 holds we make use of the fact that H is a normal subgroup in G. Let $A_a=\binom{0,\cdots}{a,\cdots}$, $A_b=\binom{0,\cdots,a,\cdots}{b,\cdots,a+b,\cdots}$, $M_m=\binom{0,1,a,b,\cdots}{0,m,am,bm,\cdots}$. Here $M_m^{-1}A_aM_m=\binom{0,\cdots}{am,\cdots}=A_{am}$ since all letters are displaced. Also $M_m^{-1}A_bM_m=A_{bm}$. Hence $A_{(a+b)m}=M_m^{-1}A_{a+b}M_m=M_m^{-1}A_aA_bM_m=M_m^{-1}A_aM_m\cdot M_m^{-1}A_bM_m=A_{am}A_{bm}=A_{am+bm}$, whence $(a+b)m=am+bm$. The other distributive law $a(r+s)=ar+as$ will not in general be true. The law E3 is an immediate consequence of property (5) of Theorem 5.4 and the fact that $x\cdot m\circ b=xm+b$ in this theorem.

We may modify the notation for points and lines in π^* and adjcin L_∞ and the points on it to obtain a pseudo-homogeneous representation for the projective plane π. Let $P^*=(x,\ y)$ in π^* be written $P=(x,\ -y,\ 1)$ in π. If the line $y=xm+b$ in π^* cuts L_∞ in π in the point Q, write $Q=(1,\ -m,\ 0)$ in π. Finally write $A=(0,\ 1,\ 0)$. Then the lines $y=xm+b$ in π^* become $y+xm+zb=0$ containing points (x, y, z) in π. Also write $x=c$ in π^* as $x+z(-c)=0$ in π, and finally L_∞ is $z=0$. The associative law of multiplication and D1 permit us to take any equation $ya+xb+zc=0$ as the equation of a line, but without the distributive law D2: $a(b+c)=ab+ac$ we cannot write the points in homogeneous form. The associative law of multiplication, M1, is not necessary in order that a ring satisfying the remaining axioms of Theorem 5.7 define a

plane, and such rings and planes will be investigated below. It is to include such rings and planes in the above transition from an affine to a projective plane that the apparently unnecessary step of changing the sign of the y-coordinate is made, since from the remaining axioms it does not follow that $-xm=x(-m)$ although it is true that $-xm=(-x)m$.

Veblen and Wedderburn([9]) have constructed certain number systems suitable for the coordinatization of planes very closely related to the rings characterized in Theorem 5.7. A more thorough investigation of these number systems and the planes which they define is given here. In §6 it will be shown that these systems correspond to a particular configuration theorem and further properties of these planes will be considered.

THEOREM 5.8. *Given a ring R in which addition and multiplication satisfy*:

A1. $(a+b)+c=a+(b+c)$,
A2. $b+a=a+b$,
A3. $a+0=0+a=a$,
A4. $a+(-a)=(-a)+a=0$,
(5.4) M5. $0a=a0=0$,
D1. $(a+b)c=ac+bc$,
E1. $xa=b$, $a\neq 0$, *has a unique solution* x,
E2. $az=b$, $a\neq 0$, *has a unique solution* z,
E3. $xr=xs+t$, $r\neq s$, *has a unique solution* x.

Then if $\phi\neq 0$ and $\psi\neq 0$ are fixed but arbitrary in R we may define points (x, y, z) of forms (x, y, ϕ), $(\phi, y, 0)$, $(0, \phi, 0)$ where x and y take all values in R. Lines are defined as those sets of points satisfying one of the equations $y\psi+xb+zc=0$, $x\psi+zc=0$, $z\psi=0$, where b and c take all values in R. Points and lines so defined form a projective plane([10]) π.

The plane π is independent of the particular choice of ϕ and ψ. We may define a new multiplication $\circ$ by the rule

$$u = v \circ s \text{ if and only if } s\psi = \phi\lambda, \qquad u\psi = v\lambda. \tag{5.5}$$

The operations $+$, $\circ$, define a ring R_1 satisfying all the properties which R does and determining the same plane. R_1 has the additional property that ϕ is a unit:

$$M3 \quad \phi \circ a = a \circ \phi = a. \tag{5.6}$$

Proof. We must first show that there is one and only one line through two distinct points P and Q. We must treat cases separately according to the

([9]) O. Veblen and J. H. M. Wedderburn [1].

([10]) The systems used by O. Veblen and J. H. M. Wedderburn are slightly different. Here the right distributive law is used rather than the left, and the roles of the x and the y are reversed. To go from their planes to these given here we consider the rings anti-isomorphic to theirs, take the dual plane, and then interchange the x and y. They do not state the analogue of the necessary condition E3, nor do they observe that the plane is independent of the choice of ϕ and ψ.

different forms for P and Q. Every case goes through almost trivially except for $P=(x_1, y_1, \phi)$, $Q=(x_2, y_2, \phi)$, and here with $x_1 \neq x_2$, $y_1 \neq y_2$. Neither P nor Q lies on $z\psi=0$ and we cannot have $x_1\psi+zc=0$ and $x_2\psi+zc=0$ with $x_1 \neq x_2$. Hence the line PQ, if there is one, is of the form $y\psi+xb+zc=0$. Now observe from D1, putting $b=-a$, that $0=0m=am+(-a)m$ whence $-am=(-a)m$. Hence if $y_1\psi+x_1b+\phi c=0$ and $y_2\psi+x_2b+\phi c=0$, by subtraction and D1 $(y_1-y_2)\psi+(x_1-x_2)b=0$ whence $(x_1-x_2)b=(-y_1+y_2)\psi$. From E1 this determines a unique b. Then from E2, $\phi c=-y_1\psi-x_1b$ determines c uniquely. From this determination of b and c we find a unique line containing P and Q. To show that two different lines intersect in one and only one point we must again treat several cases according to the form of the lines. These all go through quite easily except when they are of the forms $y\psi+xb_1+zc_1=0$, $y\psi+xb_2+zc_2=0$, and here only when $b_1 \neq b_2$. Since $\psi \neq 0$ it cannot be a point $(0, \phi, 0)$ or $(\phi, y, 0)$ with $b_1 \neq b_2$. Hence it must be a point (x, y, ϕ). Substituting $z=\phi$ we have $xb_1=xb_2+\phi c_2-\phi c_1$ and by E3, since $b_1 \neq b_2$, there is a unique solution x. Then $y\psi=-xb_1-zc_1$ by E1 determines y uniquely and we find a unique point (x, y, ϕ) on both lines. Hence these points and lines form a projective plane π.

To show that π is independent of the particular choice of ϕ and ψ, we show first that for ϕ fixed it is independent of ψ, and then that for ψ fixed it is independent of ϕ.

Let π_1 be the plane determined by ϕ and ψ_1, and π_2 be the plane determined by ϕ and ψ_2. The relation $u\psi_1=u'\psi_2$ determines by E1 a 1-1 mapping $u \rightleftarrows u'$ of R into itself leaving 0 fixed. Also let $\phi c=-a\psi_1$, $\phi d=-a\psi_2$ determine the mapping $c \rightleftarrows d$. The correspondences

$$
\begin{aligned}
(x, y, \phi) &\rightleftarrows (x, y', \phi),\\
(\phi, y, 0) &\rightleftarrows (\phi, y', 0),\\
(0, \phi, 0) &\rightleftarrows (0, \phi, 0),\\
y\psi_1 + xb + zc = 0 &\rightleftarrows y\psi_2 + xb + zc = 0,\\
x\psi_1 + zc = 0 &\rightleftarrows x\psi_2 + zd = 0,\\
z\psi_1 = 0 &\rightleftarrows z\psi_2 = 0,
\end{aligned}
\tag{5.7}
$$

set up an isomorphism between π_1 and π_2. Hence π is independent of the choice of ψ.

Now let π_1 be the plane determined by ϕ_1 and ψ and π_2 be determined by ϕ_2 and ψ. By E2, $\phi_1u=\phi_2u'$ determines a 1-1 mapping of R onto itself leaving 0 fixed. Also let $y \rightleftarrows \bar{y}$ be determined by $y\psi=-\phi_1b$, $-\phi_2b=\bar{y}\psi$. Then the correspondences

$$
\begin{aligned}
(x, y, \phi_1) &\rightleftarrows (x, y, \phi_2),\\
(\phi_1, y, 0) &\rightleftarrows (\phi_2, \bar{y}, 0),\\
(0, \phi_1, 0) &\rightleftarrows (0, \phi_2, 0),
\end{aligned}
\tag{5.8}
$$

$$
\begin{gathered}
y\psi + xb + zc = 0 \rightleftarrows y\psi + xb + zc' = 0, \\
x\psi + zc = 0 \rightleftarrows x\psi + zc' = 0, \\
z\psi = 0 \rightleftarrows z\psi = 0,
\end{gathered} \tag{5.8}
$$

are an isomorphism between π_1 and π_2.

Now let us form a new ring R_1 whose elements are the same as those of R. In R_1 addition is the same as in R, but multiplication is a new operation ($\circ$) as defined by (5.5), using arbitrary but fixed values for ϕ and ψ. In R_1 A1, 2, 3, 4 are clearly satisfied since the operation of addition is the same as in R. M5 for R_1 comes immediately from (5.5). Let $a \circ m = w$, $b \circ m = z$, whence, by (5.5), $m\psi = \phi c$, $a\psi = wc$, $b\psi = zc$. Adding the last two of these and using D1 for R we have $(a+b)\psi = (w+z)c$, whence, by (5.5), $(a+b) \circ m = w+z = a \circ m + b \circ m$, proving D1 for R_1. Now $x \circ a = b$ with $a \neq 0$ is equivalent to $a\psi = \phi c$ and $b\psi = xc$, where the first determines $c \neq 0$ uniquely and the second determines x uniquely, proving E1 for R_1. Similarly E2 for R_1 follows immediately. If $x \circ r = x \circ s + t$, $r \neq s$, we have with $x \circ r = z$, $x \circ s = w$,

$$
\begin{gathered}
z = w + t, \qquad r\psi = \phi m, \qquad s\psi = \phi n, \\
z\psi = xm, \qquad w\psi = xn, \qquad z\psi = w\psi + t\psi, \qquad xm = xn + t\psi.
\end{gathered} \tag{5.9}
$$

Since $r \neq s$, m is different from n, and so by E3 for R the last equation determines x uniquely. This proves property E3 for R_1. This shows that R_1 satisfies all the properties which R does. Also in R_1 if $a \circ \phi = b$, then $\phi\psi = \phi m$, $b\psi = am$, whence $m = \psi$ and $b = a$. If $\phi \circ a = b$ then $a\psi = \phi m$, $b\psi = \phi m$, whence $a\psi = b\psi$ and $a = b$. Hence in R_1 the rule M3, $\phi \circ a = a \circ \phi = a$, is satisfied.

It remains to show that R_1 defines the same plane π which R defines. For this purpose we shall take ϕ and ψ from R to determine the plane and ϕ and ϕ from R_1. Remembering that in R_1 we may write ϕ as 1, we see that the isomorphism is

$$
\begin{gathered}
R \rightleftarrows R_1, \\
(x, y, \phi) \rightleftarrows (x, y, 1), \\
(\phi, y, 0) \rightleftarrows (1, y, 0), \\
(0, \phi, 0) \rightleftarrows (0, 1, 0), \\
y\psi + xb + zc = 0 \rightleftarrows y + x \circ b' + z \circ c' = 0, \\
x\psi + zc = 0 \rightleftarrows x + z \circ c' = 0, \\
z\psi = 0 \rightleftarrows z = 0,
\end{gathered} \tag{5.10}
$$

where $c'\psi = \phi c$, $b'\psi = \phi b$. Note that the isomorphism does not alter the coordinates of points. If we use the correspondence $s \rightleftarrows s'$, where $s'\psi = \phi s$, then from (5.5) we have immediately $(v \circ s')\psi = vs$. Multiplying the line equations for R_1 on the right by ψ and using this rule, we see immediately that they are equivalent to the line equations from R just as the correspondence is given in (5.10).

This process is of course reversible in going from line equations from R to those from R_1. This establishes the isomorphism, and completes the proof.

THEOREM 5.9. *If a plane π is determined by a ring R satisfying axioms* A1, 2, 3, 4, M3, 5, D1, *and* E1, 2, 3 *as in Theorem* 5.8 *then R is the ternary ring determined by π as in Theorem* 5.4, *where $x \cdot m \circ b = xm + b$, taking A, B,* (0, 0, 1) *and* (0, −1, 1) *as the fundamental quadrilateral. The correspondence between the two representations is given by*

$$
\begin{aligned}
A &\rightleftarrows A,\\
B &\rightleftarrows B,\\
(0,0) &\rightleftarrows (0,0,1),\\
(1,1) &\rightleftarrows (1,-1,1),\\
x = c &\rightleftarrows x + (-c) = 0,\\
y = xm + b &\rightleftarrows y + xm + b = 0.
\end{aligned}
\tag{5.11}
$$

The mapping $(x, y) \to (x+s, y+t)$ for arbitrary s, $t \subset R$ is a collineation of π leaving the points of L_∞ fixed. If a plane π_1 is determined by such a ring R_1 and π_2 is any plane isomorphic to π_1 with $A_1 \rightleftarrows A_2$, $B_1 \rightleftarrows B_2$ then in the ring R_2 of π_2, $x \cdot m \circ b = x \circ m + b$ where the addition (+) *and multiplication* (∘) *satisfy axioms* A1, 2, 3, 4, M3, 5, D1, *and* E1, 2, 3. *Moreover R_2 may be considered as containing the same elements as R_1 where addition is the same and multiplication* (∘) *may be determined from that of R_1 by the rule*

$$
w = x \circ d \text{ if and only if } ds = at,\ ws = xt, \tag{5.12}
$$

for appropriate fixed $a \neq 0$, $s \neq 0$, and all x and d.

Proof. Given a ring R with unit determining a plane as in Theorem 5.8. The correspondences (5.11) determine points (x, y) and lines $x = c$, $y = xm + b$ satisfying the postulates for a Euclidean plane isomorphic to the finite part of the projective plane given by the pseudo-homogeneous coordinatization from R. Let us now introduce coordinates in this plane according to the rules of Theorem 5.4 with A, B, (0, 0) and (1, 1) as the fundamental quadrilateral. A line $y = xm$ goes through (0, 0) and $(1, m)$ and hence is the line $y = x \cdot m \circ 0$. The line $y = xm + b$ with $b \neq 0$ is parallel to $y = xm$ and goes through $(0, b)$ and hence is $y = x \cdot m \circ b$. Here lines $x = c$ are clearly the same in both representations. Hence $x \cdot m \circ b = xm + b$ and R is the ring determined by π as in Theorem 5.4.

The mapping $(x, y) \to (x+s, y+t)$ takes the points of the line $x = c$ into those of $x = c + s$ and the points of $y = xm + b$ into those of $y = xm + (sm + b - t)$ whence it is a collineation leaving the slopes of lines unchanged and hence leaving the points of L_∞ fixed. As s and t are arbitrary these collineations are simply transitive on the finite points of π.

Given a plane π_1 determined by a ring R_1, let π_2 be a plane isomorphic

to π_1 with $A_1 \rightleftarrows A_2$, $B_1 \rightleftarrows B_2$. Since there are collineations of π_1 transitive on its finite points and leaving A and B fixed there will be such an isomorphism which also maps (0, 0) of π_1 onto (0, 0) of π_2. Hence for an appropriate s, $y=xs$ of π_1 is mapped onto $y=x$ of π_2, and we have $(x, xs) \rightarrow (u, u)$. Now replace the symbols used in π_2 so that this mapping becomes $(x, xs) \rightarrow (x, x)$. Hence in general, mapping π_1 onto π_2,

$$(x, ys) \rightarrow (x, y). \tag{5.13}$$

Now consider the natural ring R_2 of π_2 (the ternary ring of Theorem 5.4) determined by the fundamental quadrilateral $A, B, O=(0, 0)$ and $I=(a, a)$. Note that the symbols used are taken from π_1 and that the unit 1 of π_1 need not be the unit a of π_2, although their respective origins coincide. A line $y=x \cdot a \circ c$ is the image of a line $y=xs+t$ where $(0, t) \rightarrow (0, c)$ whence $t=cs$ and a point (x, y) on it is the image of (x, ys). Hence $y=x+c$ in R_2 if and only if $ys=xs+t=xs+cs=(x+c)s$ in R_1, whence $y=x+c$ in R_1 and so addition in R_2 is the same operation as addition in R_1. In R_2, $y=x \cdot d \circ 0$ goes through $(0, 0)$ and (a, d) and so is the image of the R_1 line through $(0, 0)$ and (a, ds), which is $y=xt$ where $ds=at$ determines t. Hence if $w=x \circ d$ in R_2 then (x, w) in π_2 is the image of (x, ws) in π_1 where $ws=xt$. Hence $ds=at$, $ws=xt$ in R_1 define $w=x \circ d$ in R_2, and this is the rule (5.12). In π_2 the line $y=x \cdot m \circ b$ goes through $(0, b)$ and is parallel to $y=x \circ m$ which goes through $(0, 0)$ and (a, m). These are the images of $(0, 0)$ and (a, ms) in π_1. Hence it is the image of $y=xt$ where $at=ms$. Also $(0, bs) \rightarrow (0, b)$ and so $y=xt+bs \rightarrow y=x \cdot m \circ b$. Hence by (5.12) $y=xt+bs \rightarrow y=x \circ m+b$, and so in R_2, $x \cdot m \circ b=x \circ m+b$. We may now verify, as in Theorem 5.7, that R_2 has properties A1, 2, 3, 4, M3, 5, E1, 2, 3 where a is the unit of R_2.

THEOREM 5.10([11]). *Rings R_1 and R_2 with properties* A, M, D, E *of Theorem* 5.9 *determine planes π_1 and π_2 isomorphic with $A_1 \rightleftarrows A_2$, $B_1 \rightleftarrows B_2$ if and only if each can be obtained from the other by rule* (5.12). *If in R_1 multiplication is associative then R_2 is isomorphic to R_1 and $R_1 \rightleftarrows R_2$ by the rule* (5.14), $ua^{-1} \rightleftarrows u$, $ua^{-1}+va^{-1} \rightleftarrows u+v$, $(ua^{-1})(va^{-1}) \rightleftarrows u \circ v$.

Proof. From Theorem 5.9, R_1 and R_2 are each determined by their planes π_1 and π_2 and the choice of the fundamental quadrilateral A, B, O, I. If $A_1 \rightleftarrows A_2$, $B_1 \rightleftarrows B_2$, R_2 may be obtained according to Theorem 5.9 from R_1 with the same addition and with multiplication given by the rule (5.12). Similarly R_1 may be obtained from R_2. If in R_1 multiplication is associative, from $ds=at$, $ws=xt$ we have $t=a^{-1}ds$, $ws=xa^{-1}ds$, $w=xa^{-1}d$ whence $x \circ d=xa^{-1}d$. Using this we easily verify the correspondences (5.14). Note that if multiplication is associative, then the substitutions $x \rightarrow xm+b$, $m \neq 0$, form a doubly transitive group as in Theorem 5.6.

([11]) If R_1 does not have an associative multiplication R_2 need not be algebraically isomorphic to R_1. See appendix II for examples.

We shall summarize the results of this section briefly. For every choice of a pair of points A, B in a plane π a set S of doubly transitive permutations is determined (Theorem 5.2) and if L, M are two of these, ML^{-1} is associated with a projection in π. Conversely a doubly transitive set S of permutations satisfying the necessary conditions determines a plane π (Theorem 5.2). Every choice of a quadrilateral A, B, O, I in π determines a ternary ring R (Theorem 5.4) with a zero and a unit. If π is an analytic plane over a skew-field F then $R=F$ with $x \cdot m \circ b = xm+b$ for every choice of A, B, O, I (equations 5.2, Theorem 5.9). A plane represented by a Veblen-Wedderburn number system is independent of the choice of ϕ and ψ, and may also be represented by a Veblen-Wedderburn number system with a unit (Theorem 5.8), and this system is one of the rings R determined by π as in Theorem 5.4 (Theorem 5.9). Multiplication in this system is associative if and only if the $A-B$ permutations S form a group (Theorems 5.6, 5.7, 5.10) or a coset of a group (Theorem 5.5).

6. **Configuration theorems.** A configuration theorem is a statement of the following sort: Given points $P_1, \cdots, P_m$ and lines $L_1, \cdots, L_n$ and the information that (1) at least certain points (P) are distinct; (2) at least certain lines (L) are distinct; (3) $P_i \subset L_j$ for certain points and lines; and (4) $P_r \not\subset L_s$ for certain points and lines, then we may conclude that there are further incidences $P_u \subset L_v$.

Axiom A3 of Veblen and Young is of this type as are also the Theorems of Desargues and Pappus. It is natural to enquire what theorems of this sort are possible and what relationships exist between such theorems. Outstanding in this connection is Hessenberg's proof that the Theorem of Desargues is a consequence of the Theorem of Pappus([12]). Moufang([13]) has investigated a number of theorems which are consequences of the Desargues Theorem and in particular has shown that the Theorem of the complete quadrilateral is definitely weaker than the Theorem of Desargues. For the most part Moufang assumes not only Hilbert's axioms of connection but also the axioms of order. Most of her results are, however, true without the hypothesis of order.

In a configuration theorem, we are given a configuration C and the conclusion leads to a configuration C^* containing the points and lines of C and further incidences. In the statement of a configuration theorem above, C^* was assumed to have exactly the same points and lines as C. It is an unimportant generalization to permit C^* to contain points and lines constructed from C. If the given configuration is not open then Moufang calls the theorem conditional. If C is open and C^* is obtained by adding a single incidence to C,

([12]) G. Hessenberg [1].

([13]) R. Moufang [1–7]. She has shown in [6] that the validity of the theorem of the complete quadrilateral is equivalent to the possibility of introducing coordinates from an alternative field, and that there exist alternative fields (that is, the Cayley algebra) which are not skew-fields. Alternative fields are covered in M. Zorn [1, 2] and in R. Moufang [7].

we shall call the theorem simple. Most theorems may be treated as a combination of several simple theorems. C is free equivalent by Theorem 4.9 to some r-point (4.3) and if C^* is also open it is free equivalent to some $(r-1)$-point, since in Theorem 4.11 P and L are unchanged and I has been increased by 1. This means that any finite configuration in the plane can be generated by an $(r-1)$-point and in a restricted way. All such theorems which the author has investigated are demonstrably impossible. We shall restrict our attention to theorems where C is open and C^* is confined. For if C^* is not open then it contains a confined configuration which may be obtained by free contraction from C^* and this may be considered the consequence of the appropriate incidences in part of C.

As a further refinement in the classification of configuration theorems we distinguish between theorems in which the points and lines of C are arbitrary, apart from the mutual relationships required of them, and theorems in which certain points and lines of C are fixed. The first type we shall call a universal theorem, the second a fixed point theorem. Hilbert's formulation of the Desargues Theorem is a fixed point theorem in that he requires the axis of perspectivity always to be the line at infinity. Theorems on webs[14] are of course fixed point theorems.

The Theorem of Pappus in a plane π is equivalent to the possibility of introducing coordinates from a commutative field F. What further configuration theorems can there be in such a plane π? Given C we may take four of its generating points as a fundamental quadrilateral and introduce F as the natural ring of π determined by this quadrilateral. The other points and lines of C may be represented by values from F, including parameters representing the arbitrariness of the choice of the generators of C. The conditions (1), (2), (4) will correspond to a relation $A(u, \cdots, w) \neq 0$ where A is an appropriate rational function of the parameters $u, \cdots, w$. After we express the coordinates of the points and the equations of the lines in terms of the parameters the conclusion $P \subset L$ will determine a relation $B(u, \cdots, w) = 0$. Hence the theorem may be stated thus: $B = 0$ whenever $A \neq 0$, whence $AB = 0$ for all values for which AB is defined. If $AB = M/N$ where M and N are polynomials, then $MN = 0$ for all values of the parameters. Here MN is a polynomial, simplified only according to the general rules of fields and not according to special properties of F, such as finite characteristic or particular relations between fixed elements of F. If MN is formally zero, then the theorem is a consequence of the field properties of F and hence of the Theorem of Pappus. If MN is not formally zero and involves no parameters, then the theorem either implies that F has finite characteristic or expresses a relation on fixed elements of F.

[14] W. Blaschke and G. Bol [1] is an excellent work on the subject of webs and includes an extensive bibliography. A long series of papers on *Topologische Fragen der Differentialgeometrie*, beginning in 1927, mostly appearing in the Abh. Math. Sem. Hamburgischen Univ., are largely concerned with web configurations.

In the latter case the entire configuration is fixed. If MN contains parameters $x_1, \cdots, x_r$ then $MN = a_0(x_1, \cdots, x_{r-1})x_r^t + \cdots + a_t(x_1, \cdots, x_{r-1})$. In this case F must be a finite field. For either $a_0(x_1, \cdots, x_{r-1}) = 0$ for all $x_1, \cdots, x_{r-1}$ of F or there is a polynomial $a_0x_r^t + \cdots + a_t$, which is not identically zero, which vanishes for all $x_r \subset F$. In either event we are led eventually to a non-zero polynomial in one argument x which vanishes for all choices of x from F. But such a polynomial has only a finite number of roots and hence F must be a finite field. In the finite field with $n = p^s$ elements every element satisfies $x^n - x = 0$. Our conclusion will be stated as a theorem:

THEOREM 6.1. *If the Theorem of Pappus is valid in a plane π, then any configuration theorem which does not consist entirely of fixed points is either* (1) *a consequence of the Theorem of Pappus,* (2) *equivalent to characteristic p for F, or* (3) *equivalent to finiteness for F and hence for π.*

In a plane π if the Desargues Theorem is valid but not the Pappus Theorem, then corresponding to any configuration theorem in π not a consequence of the Desargues Theorem, there is an identity in the coordinatizing skew-field K. A universal theorem in π will correspond to an identity expressible entirely in terms of the unit 1 and parameters. Here there is no canonical form for rational functions of more than one indeterminate and little is known about the possibility of universal identities[15]. If there is a one-parameter universal identity in K, it can, as above, be shown to be equivalent to a polynomial identity with coefficients from the characteristic subfield. Hence, as $0, 1, \cdots$ must be roots, the characteristic must be a finite p. If this identity is $F(x) = 0$ where $F(x)$ is a polynomial which does not vanish identically, then we may suppose without loss of generality that $F(x)$ has no multiple factors. In this case if n is the least common multiple of the degrees of the divisors of $F(x)$ then, from well known theorems on polynomial decomposition modulo p, $F(x)$ is a divisor of $x^N - x$, where $N = p^n$. Hence $x^N = x$ for every x in K, and $x^{N-1} = 1$ for every $x \neq 0$ in K. For $x \neq 0$, $x^{-1} = x^{N-2}$ and so the subfield generated by elements $x, y, \cdots, z$ is a polynomial ring in these elements. Consider the subfield $Q(x, y)$. Every element in Q is a linear combination with coefficients $0, 1, \cdots, p-1$ of elements from the group $G(x, y)$ generated by x and y. Every element in G has order dividing $N-1$. If Burnside's conjecture is correct G must be finite[16]. If this conjecture is true, then Q must be finite, and by the Wedderburn Theorem Q is a Galois field and $yx = xy$. Hence K is commutative, whence K itself is a finite field containing at most

[15] In D. E. Littlewood [1] and A. R. Richardson [1] polynomial relations with coefficients from K are considered. W. Wagner [1] considers universal polynomial relations and shows that none exist in ordered planes, but that there always exist universal identities in division algebras, since the finite basis permits a matric representation and matric identities exist. B. H. Neumann [1] and others have treated the analogous problem for groups.

[16] This conjecture is discussed in B. H. Neumann [1]. Analogous ring and field problems are discussed in A. Kurosch [1].

p^n elements. But even if this conjecture on groups is false it may very well be true that $x^N=x$, $N=p^n$ for all elements of a skew-field K, implies that K is finite (and hence is a Galois field with at most p^n elements). For $n=1$, $F(x)=x(x-1)\cdots(x-p+1)$ and K is the characteristic field $K(p)$. For $n=2$ every element of K not in $K(p)$ satisfies an equation of second degree over $K(p)$. Hence $x^2=ax+b$, $y^2=cy+d$, and x^2 and y^2 are expressible as functions of first degree. Also $(x+y)^2=g(x+y)+h$ and $x^2+xy+yx+y^2=$1st degree, whence $yx=-xy+$1st degree. Here every third degree expression in x and y is also expressible as one of at most second degree. For example $xyx=x(-xy+$1st degree$)=-x^2y+$2nd degree$=$(1st degree$)y+$2nd degree $=$2nd degree. Hence every expression of third degree or higher can be reduced to second degree, and $1, x, y, xy$ are a basis of $Q(x, y)$ whence Q is finite and K is finite and contains at most p^2 elements.

For $n=3$ let us suppose first that $p\neq 2$. We have x^3, y^3, $(x+y)^3$, $(x-y)^3$ all of second degree. Combining these we get

$$x^2y + xyx + yx^2 + xy^2 + yxy + y^2x = \text{2nd degree},$$
$$-x^2y - xyx - yx^2 + xy^2 + yxy + y^2x = \text{2nd degree},$$

whence since $p\neq 2$

$$x^2y + xyx + yx^2 = \text{2nd degree},$$
$$xy^2 + yxy + y^2x = \text{2nd degree}.$$

Using these we need retain for a basis of Q only x^2y, xyx, xy^2, yxy of third degree, and x^2yx, x^2y^2, $xyxy$, $yxyx$ of fourth degree. Using $(xy)^3$, $(yx)^3$, $(xy+x)^3$, $(yx+y)^3$ as of fourth degree, we find that $xyxyx$ and $yxyxy$ are of fourth degree. Hence of fifth degree terms we need retain only x^2yxy and $xyxy^2$, and of sixth only x^2yxy^2, and no terms at all of seventh or higher degree. Hence Q is finite and K is finite with at most p^3 elements. A similar but slightly longer argument holds for $p^3=2^3=8$. On the basis of these computations and other considerations the author makes the conjecture that in a Desarguesian plane there can be a universal five point theorem which is not a consequence of the Desargues theorem only if the plane is finite.

A universal theorem on six or more points corresponds to a universal identity with two or more parameters. As noted above no such theorems exist in an ordered plane, but theorems certainly exist in planes coordinatized by division algebras (skew-fields with a finite basis over their center). In any quaternion algebra([17]) if x, y, z are three arbitrary elements, then $z(xy-yx)^2=(xy-yx)^2z$, or in words, the square of any commutator is in the center. This identity corresponds to a seven point universal theorem in Desarguesian planes. To find the fullest algebraic consequences of this geometric theorem, we prove the following theorem.

([17]) For the general quaternion algebra, see A. A. Albert [1, p. 145].

THEOREM 6.2. *Given a skew-field K whose center is F. If $z(xy-yx)^2=(xy-yx)^2z$ for three arbitrary elements of K, then either $K=F$ or K is a quaternion algebra over F.*

Proof. Here the square of every commutator $xy-yx$ is in F.

LEMMA 1. *An arbitrary element x of K not in F satisfies an equation $x^2=rx+s$ with $r, s\subset F$.*

The proof is relatively simple if the characteristic of K is not 2. Here let y be an element which does not permute with x and write $u=yx-xy$, $u^2=a\subset F$, $a\neq 0$, $v=yx^2-x^2y$, $v^2=ab\subset F$. Also $v=ux+xu$ and $vu=uv=ax+uxu$. Now $u+v=y(x^2+x)-(x^2+x)y$ is a commutator, whence $(u+v)^2=d\subset F$ and so $u^2+2uv+v^2=d$, and since the characteristic is not 2, $uv=c\subset F$. Also $ux=x(yx)-(yx)x$ is a commutator and so $uxux=ae\subset F$. Hence $c=ax+uxu$, $cx=ax^2+ae$, and as $a\neq 0$ we may divide by a and obtain $x^2=rx+s$, $r, s\subset F$.

When K is of characteristic 2 the above proof fails and the proof is more difficult. As before, choose a y which does not permute with x and write $yx+xy=u$, $u^2=a\neq 0$, $yx^2+x^2y=v$, $v^2=ab$, $v=ux+xu$. Again $uv=vu=ax+uxu$ and $uxux=ae\neq 0$ since ux is a commutator. Here $uvx=ax^2+ae$. Clearly x permutes with the right and hence also with the left of this equation. Since u and v permute we may square and obtain $u^2v^2x^2=a^2x^4+a^2e^2$ whence $a^2bx^2=a^2x^4+a^2e^2$ and so $x^4+bx^2+e^2=0$ is an equation of fourth degree over F which x satisfies. If $f(x)=x^4+bx^2+e^2$ has a linear factor, then b is a square and $f(x)$ is the square of a polynomial of second degree. Hence if $f(x)$ is reducible over F then x satisfies an equation of second degree over F. In particular if the choice of a different y gives different values for b and e, then x satisfies an equation of second degree over F and the lemma is proved. Hence suppose that b and e depend solely upon x and are independent of y. Now $uxux=ae=a(x^4+bx^2)/e$ whence $uxu=a(x^3+bx)/e$, $uxa=a(x^3+bx)u/e$, $ux=(x^3+bx)u/e$. From this we easily calculate $yx=xy+u$, $yx^2=xyx+ux=x^2y+[x^3+(b+e)x]u/e$, $yx^3=x^3y+(b+e)u$. Consequently $d=x^3+(b+e)x$ permutes with y. Since d clearly permutes with every element which permutes with x, and also with an arbitrary element y which does not permute with x, it follows that d is in the center F. As x satisfies an equation of degree lower than the fourth, from earlier remarks x must satisfy an equation of second degree over F. This completes the proof of the lemma.

LEMMA 2. *If $yz=zy$, then either* (1) *$1, y, z$ are linearly dependent over F or* (2) *F has characteristic 2 and $y^2=a$, $z^2=b$, with $a, b\subset F$.*

Suppose that $1, y, z, yz$ are linearly dependent over F. Then either $1, y, z$ are linearly dependent over F or $yz=cy+dz+e$. Here $(y-d)z=cy+e$, $z=(y-d)^{-1}(cy+e)=my+n$ and in either case $1, y, z$ are dependent over F.

Hence if $1, y, z, yz$ are dependent, so are $1, y, z$.

Let $y^2=my+a$, $a\neq 0$, $z^2=nz+b$, $b\neq 0$, $(yz)^2=eyz+f$, $f\neq 0$. Then $(my+a)(nz+b)=eyz+f$ and $(mn-e)yz+anz+bmy+ab-f=0$. Here either $1, y, z, yz$ are dependent and the lemma follows or $m=n=e=0$ and $ab=f$. Now let $(y+z)^2=p(y+z)+q$. Here $a+2yz+b=py+pz+q$, and if the characteristic is not 2, $1, y, z, yz$ are dependent and the lemma follows. If the characteristic is 2, then $y^2=a$, $z^2=b$, $(y+z)^2=a+b$ and no further conclusion may be drawn. This is the second alternative in the lemma.

The remainder of the proof is quite straightforward and will be merely sketched. It is easy to show that if x and y are two elements of K which do not permute then $1, x, y, xy$ are a basis over F for elements generated by x and y. As in A. A. Albert [1, p. 145] we may suppose x and y so chosen that $x^2=x+\beta$, $xy=y(1-x)$, $y^2=\gamma$, $4\beta+1\neq 0$. Let z be an arbitrary element of K. We wish to show that z belongs to the quaternion algebra $Q(1, x, y, xy)$. Replacing z by $z+nx$, if necessary, we have $z^2=c$. If z permutes with x, by Lemma 2 z is dependent on 1 and x and so $z\subset Q$. Suppose then that $zx\neq xz$. From the equation $(x+z)^2=p(x+z)+q$ we find that $(p-1)(xz-zx)=0$ whence $p=1$ and so we find that $zx+xz=z+r$, $r\subset F$. Now putting $z'=z+dx+e$ we can determine d and e so that $z'^2=c'$, $z'x+xz'=z'$. Now it is easy to verify that $(yz')x=x(yz')$ whence, by Lemma 2, $yz'=s+tx$, since with $x^2=x+\beta$ the second alternative cannot arise. Hence $z'=y^{-1}(s+tx)\subset Q$ and also $z\subset Q$. Consequently $K=Q$ and we have proved the theorem. Geometrically we have observed that in those planes coordinatized by quaternion algebras there is a seven point configuration theorem, not a consequence of the Desargues Theorem, which corresponds to the universal identity $z(xy-yx)^2=(xy-yx)^2z$. Conversely we have proved that the assumption of the Desargues Theorem and this additional seven point theorem leads either to the Theorem of Pappus (when K is commutative) or to one of the planes coordinatized by a quaternion algebra.

Although little is known about configuration theorems stronger than the Desargues Theorem, we are somewhat better informed about theorems weaker than the Desargues Theorem. The series of papers by Moufang is concerned with the theorem of the complete quadrilateral, and some special cases of this theorem and the theorems on webs, in particular the Reidemeister figure and the Thomsen triangle as given in the series of papers *Topologische Fragen der Differentialgeometrie*, are all consequences of the Theorem of Desargues. In fact the author knows of no valid universal theorem which is not at least consistent with the Theorem of Desargues. Theorems exist which are neither stronger than the Theorem of Desargues, nor consequences of it. Theorems of this sort known to the author are those which together with the Desargues Theorem would imply a finite characteristic for the coordinatizing skew-field.

A configuration of considerable interest is given in Figure 4. It is the spe-

cial case of the Desargues configuration in which the center of perspectivity lies on the axis of perspectivity. The associated configuration theorem is given below.

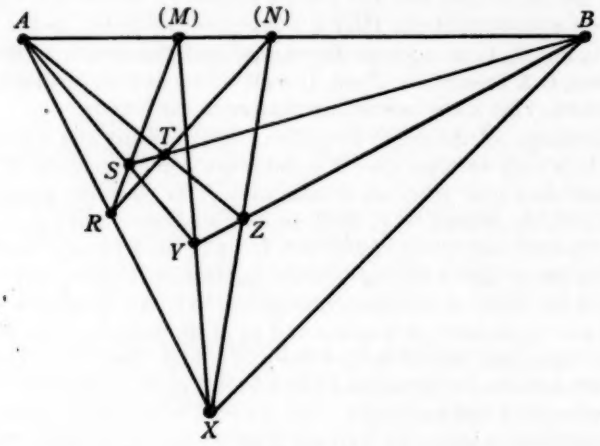


FIG. 4

THEOREM L. *Given the lines AMNB, ARX, ASY, ATZ, RSM, RTN, XYM, XZN, BYZ, then also BST is a line.*

From the same configuration several other equivalent configuration theorems may be formulated. This is clearly a 4-web theorem as it stands, involving the pencils through A, B, $M=(m)$, and $N=(n)$. Consider the ternary ring with A, B, O, I as the fundamental quadrilateral, where OI intersects AB at $N=(n)=(1)$. The points X, R on AX and Y on XM are arbitrary after A, B, M, and N are fixed. Let us take $X=(0, 0)$, $R=(0, b)$, $Y=(a, am)$. Then $S=(a, a\cdot m \circ b)$, $Z=(am, am)$, and $T=(am, am+b)$. The conclusion that BST is a line leads to

$$a\cdot m \circ b = am + b. \tag{6.1}$$

Hence Theorem L with A, B, M, N, X as fixed points is precisely equivalent to the rule (6.1) which expresses the ternary relation as a combination of two binary relations. Or, otherwise expressed, it is the rule that a line be associated with a linear equation.

Now let us keep A, B, M, N fixed and let X be arbitrary. Write $X=(a, b)$, $R=(a, c)$ where a, b, c are arbitrary. The line ASY also may be taken arbitrarily as $x=k$. The equations of the lines, using (6.1), will be $y=x+d$ for XN with $b=a+d$, $y=xm+e$ for XM with $b=am+e$, $y=x+f$ for RN with $c=a+f$, and $y=xm+g$ for RM with $c=am+g$. The points now become

$Z=(h, h+d)$, $Y=(k, h+d)$ where $h+d=km+e$ determines h, $T=(h, h+f)$, $S=(k, km+g)$. The application of the theorem now yields $km+g=h+f$. The relevant equations are

$$b = a + d, \quad b = am + e, \quad c = a + f, \quad c = am + g,$$
$$h + d = km + e, \tag{6.2}$$
$$km + g = h + f,$$

where a, b, c, k are arbitrary, the first five equations determine d, e, f, g, h and the final equation is a consequence of the theorem. Instead of taking a, b, c, k as the arbitrary elements we might equally well take a, e, g, h as the arbitrary elements and consider the system of equations

$$a + d = am + e, \quad a + f = am + g, \quad h + d = km + e, \tag{6.3}$$
$$h + f = km + g,$$

where the last equation is a consequence of the first three. Here taking $e=0$, $h=0$ we have $a+d=am$, $a+f=am+g$, $d=km$, and $f=km+g=d+g$ whence

$$a + (d + g) = (a + d) + g \tag{6.4}$$

for arbitrary a and g, and d determined by $a+d=am$. This is a restricted associative law of addition. We may use this to prove the general associative law for addition. In (6.3) we may take d, e, g, h as arbitrary since, by (6.1) and property (5) of Theorem 5.4, if $m\neq 1$, $a+d=am+e$ determines a uniquely from d, e and m. Here take $e=0$. Then $a+d=am$, $a+f=am+g=(a+d)+g$ whence by (6.4) $f=d+g$. Here $h+d=km$ and by substitution the final equation becomes

$$(h + d) + g = h + (d + g) \tag{6.5}$$

where h, d, g are all arbitrary. Since by Theorem 5.4 there is a zero for addition and in an equation $r+s=t$ any two of r, s, t determine the third uniquely, the associative law (6.5) is all that is needed to prove that addition forms a group. It is readily verified that if addition forms a group, then the last of equations (6.2) is a consequence of the others. Hence Theorem L for A, B, M, $N=(1)=C$ fixed is equivalent to the linearity (6.1) of the ternary relation and the fact that addition is a group. Note that this does not require the addition to be an abelian group. Note in Figure 4 that permutations of M, N, B do not alter the configuration but that A plays a special role.

Let us apply Theorem L again with A as the center of the pencil, $y=$constant; B the center of the pencil, $x=$constant, $M=(m)$, $N=(1)=C$. If we take $X=(0, 0)$, $R=(a, 0)$, and $Z=(b, b)$ with a, b arbitrary we find that XN is $y=x$, XM is $y=xm$, RN is $y=x-a$, RM is $y=xm-am$. Hence $Y=(b, bm)$, $T=(c, b)$ where $b=c-a$, whence $c=b+a$, $T=(b+a, b)$. Here $S=(d, bm)$ where $bm=dm-am$, whence $bm+am=dm$. By application of

Theorem L, BST are collinear whence $d=b+a$ and we have

$$bm + am = (b + a)m \tag{6.6}$$

the right distributive law.

Let us apply Theorem L again with M as $x=$constant, N as $y=$constant, $A=(1)$, $B=(m)$. We may take $X=(0, 0)$, $R=(a, a)$ and ASY as $y=x+b$ where a and b are arbitrary. Then $S=(a, a+b)$, $Y=(0, b)$, whence BYZ is the line $y=xm+b$ and $Z=(c, 0)$ where $0=cm+b$. Then ATZ is the line $y=x-c$ and since RTN is $y=a$, $T=(a+c, a)$. By Theorem L, BST are collinear. Let its equation be $y=xm+d$ where d is determined by $a+b=am+d$. The conclusion is $a=(a+c)m+d$. Hence if $0=cm+b$ and $a+b=am+d$ then $a=(a+c)m+d$. Instead of considering a and b as the arbitrary elements we may take b and d as the arbitrary elements. Here by (6.6) and the fact that addition forms a group we have $cm=-b$, $-am+a=d-b$, $a=am+cm+d$, $-am+a=-b+d$, whence $d-b=-b+d$ and so

$$b + d = d + b \tag{6.7}$$

for arbitrary b and d, the commutative law of addition. Hence the natural ring R of the plane π satisfies all the conditions of Theorem 5.8 and by Theorem 5.9 this ring is a Veblen-Wedderburn number system with a unit.

THEOREM 6.3. *If configuration Theorem* L *is satisfied in a plane π with L_∞ fixed for all choices of A, B, M, N on L_∞ with three of these as the center of the pencil $x=$constant, the center of the pencil $y=$constant, and the point* (1), *then R, the natural ring determined by a fundamental quadrilateral A (center of $x=$constant), B (center of $y=$constant), $O(0, 0)$, and $I(1, 1)$, is a Veblen-Wedderburn number system with unit. Conversely if π is a plane determined by a Veblen-Wedderburn system, then configuration Theorem* L *is valid for all choices of A, B, M, N on L_∞.*

Proof. In establishing rules (6.1) to (6.7) we have used Theorem L only with A, B, M, N as including the three points A, B, $C=(1)$ determined by a fundamental quadrilateral A, B, O, I. This proves the first part of the theorem.

The converse part of the theorem is stronger than the first part, in that the points A, B, M, N of the configuration, apart from lying on L_∞, need bear no special relation to the coordinatizing fundamental quadrilateral A, B, O, I. As noted in Theorem 5.9 the translations $(x, y)\rightarrow(x+s, y+t)$ are collineations of π leaving the points of L_∞ fixed. Hence without loss of generality we may take $X=(0, 0)$ in verifying the validity of Theorem L. Suppose $A=(p)$, $B=(q)$, $M=(m)$, $N=(n)$ is a choice of four distinct points on L_∞. Here $R=(a, ap)$ with a arbitrary and $Y=(b, bm)$ with b arbitrary. Line equations are as follows: $XRA: y=xp$, $XYM: y=xm$, $XZN: y=xn$, $RSM: y=xm-am+ap$, $RTN: y=xn-an+ap$, $BYZ: y=xq-bq+bm$ and $ASY: y=xp-bp$

$+bm$. Here if $S=(c, d)$ and $Z=(e, f)$ we have

$$(6.8)\quad d = cm - am + ap, \quad d = cp - bp + bm, \quad f = en, \quad f = eq - bq + bm.$$

From the first two of these we have $(c-a-b)m=(c-a-b)p$ whence, since $m\neq p$, $c-a-b=0$ and $c=a+b$, whence $d=bm+ap$, $S=(a+b,\ bm+ap)$. Here the equation of ATZ is $y=xp-ep+en$. Hence if T is (g, h) we have $h=gn-an+ap$, $h=gp-ep+en$ and, solving as for c and d, $g=a+e$, $h=ap+en$, $T=(a+e,\ ap+en)$. It remains to show that BST are collinear. The equation of BS is $y=xq-aq-bq+bm+ap$. Substituting the coordinates of T we must verify that $ap+en=(a+e)q-aq-bq+bm+ap$. Here substituting for en from (6.8) we have $ap+eq-bq+bm=aq+eq-aq-bq+bm+ap$, which is a true identity and so BST are collinear.

Since the center of the pencil $y=$constant may be represented by (0), this verifies the validity of Theorem L for all choices of A, B, M, N on L_∞ except those including the center of the pencil $x=$constant as one of the points. This requires two additional verifications, both similar to that given above but somewhat simpler.

This is an affine theorem in that L_∞ is fixed but no points on L_∞ are fixed. Note that for every choice of fundamental quadrilateral A, B, O, I with A and B on L_∞ we get a Veblen-Wedderburn number system, but that these systems need not be algebraically isomorphic. For two different choices of the fundamental quadrilateral it will be found (providing we alter the coordinates suitably) that addition is the same but that one multiplication may be found from the other by a rule on the order of (5.12), which is the special rule when $A_1=A_2$, $B_1=B_2$. In appendix II are examples of non-isomorphic systems defining the same plane. The number of non-isomorphic systems is clearly connected with the nature of the collineation group of the plane. The collineation group need not be doubly transitive on the points of L_∞, but in these examples and others not here given the group is simply transitive on the points of L_∞. It would be interesting to investigate the collineation groups of these planes and the geometric equivalence of the Veblen-Wedderburn systems more thoroughly.

Now let us consider those planes in which Theorem L is a universal theorem, permitting the line $ABMN$ to be arbitrary. Naturally the laws holding for the affine theorem will also hold for the universal theorem. Let us first take $A=A$, $B=O$, $M=(0, a)$, $N=(0, b)$, $X=B$, $R=(m)$ and let the line ASY be $x=c$. Here $a\neq 0$, $b\neq 0$, $a\neq b$, $c\neq 0$, and $m\neq 0$ are arbitrary. Here the equations of the lines are RSM: $y=xm+a$, RTN: $y=xm+b$, MYX: $y=a$, NZX: $y=b$. Here if $Y=(c, a)$ and $a=cr$, BYZ is $y=xr$, whence $Z=(d, b)$ with $b=dr$. Then S as the intersection of ASY and RSM is $S=(c, cm+a)$, while T as the intersection of ATZ and RTN is $T=(d, dm+b)$. The theorem then states that BST are collinear and so if $cm+a=cs$ determines s, then $dm+b=ds$. Take c, d, r, m instead of a, b, c, m as the arbitrary elements. The

theorem becomes: if $cm+cr=cs$ $(c\neq 0)$, then $dm+dr=ds$. Putting $c=1$ we have $m+r=s$, whence

$$dm + dr = d(m + r), \tag{6.9}$$

the left distributive law. Clearly this law is a consequence of the left distributive law and so is equivalent to it. The two distributive laws have as immediate consequences

$$-am = (-a)m = a(-m). \tag{6.10}$$

Finally let us apply Theorem L with $A=(0, 0)$, $B=B$, $M=(a, 0)$, $N=(b, 0)$, $X=A$, $R=(0, c)$ and $Y=(a, am)$ where $a\neq 0$, $b\neq 0$, $a\neq b$, $c\neq 0$, $m\neq 0$ are arbitrary. Here $Z=(b, am)$. Line equations will be: $ABMN$: $y=0$. ARX: $x=0$, ASY: $y=xm$, ATZ: $y=xn$ where $am=bn$, MSR: $y=xp+c$ where $0=ap+c$, MXY: $x=a$, NRT: $y=xq+c$ where $0=bq+c$, NXZ: $x=b$, BYZ: $y=am$. We now find $S=(d, dm)$ where $dm=dp+c$ and $T=(e, en)$ where $en=eq+c$. The theorem that BST are collinear yields $en=dm$. The relevant equations are

$$\begin{aligned} &am = bn, && 0 = ap + c, && 0 = bq + c, \\ &dm = dp + c, && en = eq + c, && en = dm. \end{aligned} \tag{6.11}$$

We may eliminate c from these equations and obtain

$$am = bn, \quad ap = bq, \quad ap = dp - dm, \quad ap = eq - en, \quad en = dm, \tag{6.12}$$

where a, b, m, p are arbitrary and the last equation is a consequence of the first four. If we put $a=1$, $p=1$, $m=1-b$, we find that $bq=1$, $db=1$, $n=q-1$, $e=1$, and finally $q=d$, whence $bq=qb=1$ and we may write $q=b^{-1}$,

$$bb^{-1} = b^{-1}b = 1, \tag{6.13}$$

stating that the right and left inverses of an element are equal. (The excluded possibility $m=0$ corresponds to $b=1$, where clearly $b^{-1}=1$.) Now put $a=1$, $p=1$, $m=1-br$ where b and r are arbitrary. This leads to $q=b^{-1}$, $n=b^{-1}-r$, $d=(br)^{-1}$, $e=r^{-1}$ and finally

$$(br)^{-1} = r^{-1}b^{-1}. \tag{6.14}$$

Here the excluded possibility $m=0$ corresponds to $br=1$ which is covered by (6.13).

Again let us put $m=a^{-1}$ and choose p so that $ap=bq=q-b^{-1}$. We have immediately $n=b^{-1}$, $ap=e(q-b^{-1})$, whence $e=1$. The last equation becomes $b^{-1}=da^{-1}$ and combined with $ap=q-b^{-1}=dp-da^{-1}$ leads to $q=dp$. The system of equations now reduces to

$$ap = q - b^{-1}, \quad ap = bq, \quad q = dp, \quad b^{-1} = da^{-1}. \tag{6.15}$$

Here $b(ap)=b(q-b^{-1})=bq-1=ap-1$ whence $(1-b)(ap)=1$, $ap=(1-b)^{-1}$ and $ap-1=(1-b)^{-1}-1=(1-b)^{-1}(1-1+b)=(1-b)^{-1}b=[b^{-1}(1-b)]^{-1}=(b^{-1}-1)^{-1}$. Also $dp-da^{-1}=q-b^{-1}=ap=(1-b)^{-1}$ and

$$\begin{aligned} d(p-a^{-1}) &= (1-b)^{-1}, \\ d[a^{-1}(ap-1)] &= (1-b)^{-1}, \\ d[a^{-1}(b^{-1}-1)^{-1}] &= (1-b)^{-1}, \\ d[(b^{-1}-1)a]^{-1} &= (1-b)^{-1}, \\ [(b^{-1}-1)a]d^{-1} &= 1-b, \\ (b^{-1}a-a)d^{-1} &= 1-b, \\ (b^{-1}a)d^{-1}-ad^{-1} &= 1-b. \end{aligned}$$

But $ad^{-1}=b$ by (6.14) and (6.15) whence $(b^{-1}a)d^{-1}=1$, and so $b^{-1}=(b^{-1}a)a^{-1}$. By (6.14) we have also $b=[(b^{-1}a)a^{-1}]^{-1}=a(a^{-1}b)$. We write these two restricted associative laws in the form

$$a^{-1}(ab) = b = (ba)a^{-1}. \tag{6.16}$$

It is quite easy to show that (6.14) is a consequence of (6.13) and (6.16). A Veblen-Wedderburn system satisfying also the left distributive law (6.9), the law of unique inverse (6.13), and the weak associative laws (6.16) is an alternative field[18]. The reader may find it an interesting exercise of ingenuity to show that from these laws the last of equations (6.12) is a consequence of the first four, but it is not necessary for our purposes to do so.

THEOREM 6.4[19]. *If Theorem* L *is satisfied for all choices of* A, B, M, N *on two lines in a plane* π, *then Theorem* L *is a universal theorem in* π *and any natural ring associated with* π *satisfies*

$$\begin{array}{ll} a\cdot m\circ b=am+b, & a(b+c)=ab+ac, \\ (a+b)+c=a+(b+c), & (a+b)m=am+bm, \\ a+b=b+a, & a1=1a=a, \\ a+0=0+a=a, & aa^{-1}=a^{-1}a=1, \quad a\neq 0, \\ a+(-a)=(-a)+a=0, & a^{-1}(ab)=b=(ba)a^{-1}, \quad a\neq 0, \end{array} \tag{6.17}$$

[18] R. Moufang [6, 7], M. Zorn [1, 2].

[19] R. Moufang derives alternative fields from the theorem of the complete quadrilateral, most of the geometric part of her proof being in her third paper. Her Theorem 3a [Moufang 3, p. 762] is essentially the same as Theorem L in this paper. But her derivation assumes order and her method of proof is certainly not valid if the diagonal points of every quadrilateral are collinear. (This is essentially the geometric configuration corresponding to characteristic in the coordinatizing ring.) Curiously enough she does not seem to have observed that her Theorem 2, after renumbering vertices, is identical with her Theorem D_9 (Theorem 1), but gives proofs of their equivalence.

and is an alternative field. The collineation group of π is transitive on the triangles of π.

Proof. We have shown that if Theorem L is satisfied for all choices of A, B, M, N on L_∞ and on $x=0$, then rules (6.17) are valid for the natural ring R determined by the quadrilateral A, B, O, I. In π there exist the following collineations:

(α) Translations:

$$\begin{aligned} (x, y) &\to (x + r, y + s), \\ A &\to A, \\ (m) &\to (m), \\ x = c &\to x = c + r, \\ y = xm + b &\to y = xm + (rm + b - s). \end{aligned}$$

(β) The reflection:

$$\begin{aligned} (x, y) &\rightleftarrows (y, x), \\ A &\rightleftarrows B, \\ (m) &\rightleftarrows (m^{-1}) && (m \neq 0), \\ x = c &\rightleftarrows y = c, \\ y = xm + b &\rightleftarrows y = xm^{-1} - bm^{-1} && (m \neq 0). \end{aligned}$$

(γ)

$$\begin{aligned} A &\to A, \\ (m) &\to (m + r), \\ (x, y) &\to (x, y + xr), \\ x = c &\to x = c, \\ y = xm + b &\to y = x(m + r) + b. \end{aligned}$$

(δ) The reflection:

$$\begin{aligned} A &\rightleftarrows A, \\ (m) &\rightleftarrows (0, m), \\ (x, y) &\rightleftarrows (x^{-1}, x^{-1}y) && (x \neq 0), \\ x = 0 &\rightleftarrows L_\infty, \\ x = a &\rightleftarrows x = a^{-1} && (a \neq 0), \\ y = xm + b &\rightleftarrows y = xb + m. \end{aligned}$$

The validity of (α) has already been noted in Theorem 5.9. (β) is easily verified by observing that if $v = um + b$, $vm^{-1} = (um)m^{-1} + bm^{-1} = u + bm^{-1}$, and

$u=vm^{-1}-bm^{-1}$. For (γ) we observe that if $y=xm+b$, then by the left distributive law $y+xr=x(m+r)+b$. For (δ) the verifications are trivial except to show that from $v=um+b$, $u\neq 0$, $u^{-1}v=u^{-1}(um)+u^{-1}b=u^{-1}b+m$.

The collineation (δ) takes L_∞ into $x=0$. (α) takes $x=0$ into $x=r$, an arbitrary line through A different from L_∞. (β) takes the pencil through A into the pencil through B. (γ) takes the pencil through B into the pencil through an arbitrary (r). Hence the collineation group of π is transitive on the lines of π, and an arbitrary line may be taken into L_∞. Hence since Theorem L is valid for all choices of A, B, M, N on L_∞, by Theorem 6.3, it is a universal theorem. Hence for an arbitrary choice of the fundamental quadrilateral, Theorem L will be satisfied for $ABMN$ on the new L_∞ and the new $x=0$ and so the natural ring will satisfy the rules (6.17).

Consider any triangle PQS of π. By a collineation we may take the line PQ into L_∞. By a combination of (γ), which leaves A fixed and is transitive on the remaining points of L, and (β), which interchanges A and B, we may take two arbitrary points of L_∞ into A and B. Then (α) leaves A and B fixed and takes an arbitrary point not on L_∞ into $O(0, 0)$. Hence there is a collineation taking PQS into ABO and the collineations of π are transitive on the triangles of π.

It is not known to the author whether or not the collineations of π are transitive on quadrilaterals. If they are, an arbitrary choice of the fundamental quadrilateral will yield the same alternative field as the natural ring for π. If they are not, every natural ring will be an alternative field, but non-isomorphic alternative fields will determine the same plane. One may be obtained from another by the rule (5.12). It may be shown that the collineations of π will be transitive on quadrilaterals if and only if for every $r\neq 0$ there exists a function $\phi_r(u)$ such that

$$(6.18)\qquad \phi_r(a)+\phi_r(b)=\phi_r(a+b),\quad \phi_r(am)=\phi_r(a)\left[r^{-1}\phi_r(m)\right].$$

Obviously, when the multiplication is associative $\phi_r(u)=ru$ satisfies these conditions.

Theorem L is a special case of the Theorem of Desargues. If we assume the Theorem of Desargues as a universal theorem, then not only the rules (6.17) must be valid for the natural rings associated with the plane, but, taking the origin as the center of perspectivity and L_∞ as the axis of perspectivity, it is easily shown that multiplication is associative, and that R is a skew-field. Then the homogeneous representation of π is easily established and the collineations (5.2) show that every Desarguesian plane is determined by a unique skew-field, and hence is an analytic plane. Conversely the verification of the validity of the Desargues Theorem in an analytic plane is a straightforward and relatively simple calculation. But to prove that every Desarguesian plane is an affine plane it is simpler to rely on the methods of standard texts than to use Theorem 6.3 which depends on the weaker Theorem L.

APPENDIX I

THEOREM. *Given permutations S on a finite number n of letters such that* (a) *S is doubly transitive,* (b) *if $M, L \subset S$ and $M \neq L$, then ML^{-1} fixes at most one letter, then there are $n-1$ permutations taking a given letter i into a given letter j and* (c) *if L does not take i into j there is exactly one M taking i into j such that ML^{-1} displaces all letters.*

Denote the letters by $1, 2, \cdots, n$. Consider the permutations $T(i, j) = \binom{\cdots, i, \ k, \cdots}{\cdots, j, a_k, \cdots}$ for fixed i, j, and $k \neq i$. a_k may take any value $1, \cdots, n$ except j. Hence there are at least $n-1$ permutations $T(i, j)$ since S is doubly transitive. If two $T(i, j)$ permutations A, B have the same value for a_k, then AB^{-1} fixes both i and k and by condition (b) $A = B$. Hence there are exactly $n-1$ permutations $T(i, j)$. Suppose L does not take i into j, $L = \binom{\cdots, i, \cdots, k, \cdots}{\cdots, s, \cdots, j, \cdots}$. The $n-2$ permutations $\binom{\cdots, \ t, \cdots, i, \cdots}{\cdots, a_k, \cdots, j, \cdots}$ with $t \neq i$, k and $\binom{i}{a_i}$ from L are different from L and each other, since L does not take i into j, and if two of these were the same N then NL^{-1} would fix two letters. Call the remaining $T(i, j)$, M. Then M takes i into j and ML^{-1} displaces all letters.

If S is a set of permutations on infinitely many letters satisfying (a) and (b) then (c) may be violated in either of two ways. Either for a given L and $\binom{i}{j}$ there may be no M taking i into j such that ML^{-1} displaces all letters or there may be more than one. Both possibilities actually arise. Consider permutations on the non-negative integers $0, 1, 2, \cdots$. Suppose given a finite set $P_1 = \binom{a_{11}, \cdots, a_{1r}, \cdots}{b_{11}, \cdots, b_{1r}, \cdots}, \cdots, P_i = \binom{a_{i1}, \cdots}{b_{i1}, \cdots}, \cdots, P_s = \binom{a_{s1}, \cdots}{b_{s1}, \cdots}$ where the a's and b's are subject to the following conditions: (A) $\binom{\cdots, c, \cdots, e, \cdots}{\cdots, d, \cdots, f, \cdots}$, $c \neq e$, $d \neq f$, occurs in at most one P for any combination of c, d, e, f and every combination with $c+d+e+f \leqq N$ does occur; and (B) For each P_i, $a_{i1}, \cdots$ are distinct and include $1, 2, \cdots, N$ and $b_{i1}, \cdots$ are distinct and include $1, 2, \cdots, N$. Then it is possible to extend $P_1, \cdots, P_s$ for $N = k$ to a larger system $P_1, \cdots, P_t$ which satisfies (A) and (B) for $N = k+1$. In a countable number of steps we have then constructed a set S of permutations satisfying (a) and (b). If all $\binom{\cdots, c, \cdots, e, \cdots}{\cdots, d, \cdots, f, \cdots}$ with $c+d+e+f = k+1$ occur in $P_1, \cdots, P_s$ then no extension is needed to satisfy (A) for $N = k+1$. If one such combination has not occurred write $P_{s+1} = \binom{c, e}{d, f}$ and with a finite number of such adjunctions (A) is satisfied for $N = k+1$. Now if a given $i \leqq k+1$ has not occurred in some P as an a, adjoin to this P the combination $\binom{i}{n}$ where if n is sufficiently large (A) will not be violated, since n must satisfy only a finite number of inequalities. Similarly adjoin $\binom{n}{i}$ to any P in which $i \leqq k+1$ has not occurred as a b. If n is large enough this may be done without violating (A). Hence in a finite number of steps we may extend $P_1, \cdots, P_s$, satisfying (A) and (B) for $N = k$, to $P_1, \cdots, P_t$ satisfying (A) and (B) for $N = k+1$. Since we do not alter the letters originally in the P's but merely adjoin new letters and new P's there is no ambiguity in considering the result of a countable number

of steps, and this is a set S satisfying (a) and (b).

If we start with

$$P_1=\begin{pmatrix}0,1,2,3,4,\cdots,n,\cdots\\0,1,2,3,4,\cdots,n,\cdots\end{pmatrix},\quad P_2=\begin{pmatrix}0,1,2,3,4,\cdots,2r-1,\ 2r,\cdots\\1,2,0,4,3,\cdots,\ 2r,2r-1,\cdots\end{pmatrix},\quad P_3=\begin{pmatrix}0,1,2,3,\cdots,\ 2r,2r+1,\cdots\\1,0,3,2,\cdots,2r+1,\ 2r,\cdots\end{pmatrix},$$

we have (A) and (B) satisfied for $k=2$. As $k\to\infty$, P_1, P_2, P_3 remain unchanged, and the complete set S satisfies (a) and (b). But here both P_2 and P_3 take 0 into 1 and $P_2P_1^{-1}$ and $P_3P_1^{-1}$ displace all letters, thus violating (c). On the other hand if we start with $P_1=\begin{pmatrix}0,1,2,\cdots,n,\cdots\\0,1,2,\cdots,n,\cdots\end{pmatrix}$, $P_2=\begin{pmatrix}0,1,2\\0,1,2\end{pmatrix}$ we have (A) and (B) satisfied for $k=2$. At each stage if some P_i does not fix any letter we may adjoin $\begin{pmatrix}c\\c\end{pmatrix}$ to P_i with a sufficiently large c. In this case there is no P such that PP_1^{-1} displaces all letters and (c) is violated the other way.

Appendix II

Of the Veblen-Wedderburn systems with nine elements, there are, besides the finite field, four non-isomorphic systems. In all four the additive group is the elementary abelian group of order 9 and the elements may be represented as 0, 1, 2, a, $a+1$, $a+2$, $2a$, $2a+1$, $2a+2$, where we add elements $ia+j$ by taking i and j modulo 3. The multiplication tables of the four systems are given below. Since $0x=x0=0$, 0 is omitted from the table. The product uv is in the row beginning with u and in the column under v. The element 2 is associative on the left since $(2x)y=(x+x)y=xy+xy=2(xy)$, but need not be in the center (System U).

R. The group system

1	2	a	$a+1$	$a+2$	$2a$	$2a+1$	$2a+2$
2	1	$2a$	$2a+2$	$2a+1$	a	$a+2$	$a+1$
a	$2a$	2	$2a+1$	$a+1$	1	$2a+2$	$a+2$
$a+1$	$2a+2$	$a+2$	2	$2a$	$2a+1$	a	1
$a+2$	$2a+1$	$2a+2$	a	2	$a+1$	1	$2a$
$2a$	a	1	$a+2$	$2a+2$	2	$a+1$	$2a+1$
$2a+1$	$a+2$	$a+1$	$2a$	1	$2a+2$	2	a
$2a+2$	$a+1$	$2a+1$	1	a	$a+2$	$2a$	2

S. $x^2=x+1$

1	2	a	$a+1$	$a+2$	$2a$	$2a+1$	$2a+2$
2	1	$2a$	$2a+2$	$2a+1$	a	$a+2$	$a+1$
a	$2a$	$a+1$	1	$2a+2$	$a+2$	2	$2a+1$
$a+1$	$2a+2$	$2a+1$	$a+2$	1	2	$2a$	a
$a+2$	$2a+1$	1	$2a$	a	$2a+2$	$a+1$	2
$2a$	a	$2a+2$	2	$a+1$	$2a+1$	1	$a+2$
$2a+1$	$a+2$	2	a	$2a$	$a+1$	$2a+2$	1
$2a+2$	$a+1$	$a+2$	$2a+1$	2	1	a	$2a$

T. $z^2=2z+1$

1	2	a	$a+1$	$a+2$	$2a$	$2a+1$	$2a+2$
2	1	$2a$	$2a+2$	$2a+1$	a	$a+2$	$a+1$
a	$2a$	$2a+1$	$a+2$	1	$2a+2$	$a+1$	2
$a+1$	$2a+2$	1	$2a$	a	$a+2$	2	$2a+1$
$a+2$	$2a+1$	$a+1$	1	$2a+2$	2	$2a$	a
$2a$	a	$a+2$	$2a+1$	2	$a+1$	$2a+2$	1
$2a+1$	$a+2$	$2a+2$	2	$a+1$	1	a	$2a$
$2a+2$	$a+1$	2	a	$2a$	$2a+1$	1	$a+2$

U. 2 not in center

1	2	a	$a+1$	$a+2$	$2a$	$2a+1$	$2a+2$
2	1	$2a$	$2a+2$	$2a+1$	a	$a+2$	$a+1$
a	$2a+1$	$a+2$	1	$2a$	$a+1$	$2a+2$	2
$a+1$	$2a$	$2a+2$	$a+2$	2	1	a	$2a+1$
$a+2$	$2a+2$	2	$2a$	$a+1$	$2a+1$	1	a
$2a$	$a+2$	$2a+1$	2	a	$2a+2$	$a+1$	1
$2a+1$	$a+1$	1	a	$2a+2$	$a+2$	2	$2a$
$2a+2$	a	$a+1$	$2a+1$	1	2	$2a$	$a+2$

System R is a group system which may be derived from a doubly transitive group on 9 letters in which only the identity fixes two letters, as in Theorems 5.6 and 5.7. The subgroup of elements fixing one letter (the multiplicative group of the Veblen-Wedderburn system) is the quaternion group.

In systems S and T the elements 0, 1, 2 are in the center and in S every other element z satisfies $z^2=z+1$, while in T every other element z satisfies $z^2=2z+1$. In system U the center consists solely of the elements 0, 1. In S, T, U it is immediately observed that $(aa)a\neq a(aa)$.

Systems S and T suggest a method of constructing Veblen-Wedderburn systems J which turns out to be valid. Given a commutative field F and a polynomial with coefficients from F, z^2-rz-s, irreducible over F. Let the elements of J be $au+b$, a, $b\subset F$, and let addition in J be defined by $(au+b)+(cu+d)=(a+c)u+(b+d)$. For multiplication let $c(au+b)=(au+b)c=acu+bc$ and $c(wv)=(cw)v$ if a, b, $c\subset F$. Also let $z^2=rz+s$ if $z=au+b$, $a\neq 0$. These rules, together with the right distributive law, completely determine the multiplication table for J. It may be verified with a certain amount of tedious calculation that J then satisfies all requirements of Theorem 5.8 and is a Veblen-Wedderburn system. J will not be a field or even a skew-field unless F is the finite field with two elements.

Although the four systems R, S, T, U are not algebraically isomorphic, they all define the same projective plane. Systems S, T, U define planes $\pi(S)$, $\pi(T)$, $\pi(U)$ isomorphic to each other with $A\rightarrow A$, $B\rightarrow B$, and each may be obtained from the other by an appropriate application of rule (5.12). To

map $\pi(T)$ onto $\pi(S)$ take $a=1$, $s=2$ in (5.12). $\pi(T)\to\pi(S)$, $(x, 2y)\to(x, y)$. The equation $(x \circ d)2 = x(d2)$ defines a product $x \circ d$ in S in terms of the multiplication in T. To map $\pi(U)$ onto $\pi(S)$, take $a=1$, $s=2a+2$ in (5.12). $\pi(U)\to\pi(S)$, $[x, y(2a+2)]\to[x, y]$. The equation $(x \circ d)(2a+2) = x[d(2a+2)]$ defines a product $x \circ d$ in S in terms of the multiplication in U. There is no isomorphism between $\pi(R)$ and $\pi(S)$ leaving A and B fixed. To map $\pi(S)$ onto $\pi(R)$ let us take the quadrilateral A, (2), (0, 0), (1, 0) in $\pi(S)$ and use it as a fundamental quadrilateral to coordinatize the plane. By Theorem 6.3 this will lead to a Veblen-Wedderburn system and the lines in the plane will be determined by equations $x=c$, $y=xm+b$. Keeping x coordinates unchanged this leads to $(x, y)\to(x, x+y)$, $(m)\to(1+m)$. In the new ring addition is found to be the same as in S, and multiplication $u \circ m$ is given by the rule $u \circ m = u(m-1)+u$. The new ring is the system R and so we have mapped $\pi(S)$ onto $\pi(R)$. Hence we have shown that all four systems define the same plane π.

What is the collineation group of π? If there were a collineation of π taking L_∞ into another line then Theorem L would be satisfied for all A, B, M, N on two lines, and hence by Theorem 6.4 every natural ring of π would be an alternative field. Since this is not the case, the collineations of π leave L_∞ fixed. Taking π as represented by R, the following mappings of points are collineations:

$$
(1)\qquad
\begin{aligned}
A &\to (2),\\
B = (0) &\to (1),\\
(1) &\to (0),\\
(2) &\to A,\\
(m) &\to (2m) \qquad (m \neq 0, 1, 2),\\
(x, y) &\to (x + y, x - y),
\end{aligned}
$$

$$
(2)\qquad
\begin{aligned}
A &\to A,\\
(m) &\to (r^{-1}ms) \qquad (r, s \text{ fixed parameters} \neq 0),\\
(x, y) &\to (xr, ys),
\end{aligned}
$$

$$
(3)\qquad
\begin{aligned}
A &\to A,\\
(m) &\to (m),\\
(x, y) &\to (x + r, y + s) \qquad (r, s \text{ fixed parameters}).
\end{aligned}
$$

These generate the complete collineation group of π which is of order $10\cdot8\cdot6\cdot4\cdot2\cdot81 = 311{,}040$. The collineations induce a transitive permutation group on the points of L_∞ of order $10\cdot8\cdot6\cdot4\cdot2 = 3840$. In the symmetric group on the ten points of L_∞, A, (0), (1), (2), (a), $(a+1)$, $(a+2)$, $(2a)$, $(2a+1)$,

$(2a+2)$, this group is the centralizer of the element $[A, (0)][(1), (2)][(a), (2a)][(a+1), (2a+2)][(a+2), (2a+1)]$.

The full projective group of π is the symmetric group on the points of a line, and hence is of relatively little interest.

Bibliography

A. A. Albert

1. *Structure of algebras*, Amer. Math. Soc. Colloquium Publications, vol. 24.

Garrett Birkhoff

1. *Combinatorial relations in projective geometries*, Ann. of Math. vol. 36 (1935) pp. 743–748.

W. Blaschke and G. Bol

1. *Geometrie der Gewebe*, Berlin, Springer, 1938.

R. D. Carmichael

1. *Groups of finite order*, Ginn and Co., 1937.

E. W. Davis

1. *A geometric picture of the fifteen schoolgirls problem*, Ann. of Math. vol. 2 pp. 156–157.

O. Eckenstein

1. *Bibliography of Kirkman's schoolgirl problem*, Messenger of Mathematics vol. 41 (1911) pp. 33–36.

A. Emch

1. *Triple and multiple systems, their geometric configurations and groups*, Trans. Amer. Math. Soc. vol. 31 (1929) pp. 25–42.

G. Hessenberg

1. *Beweis des Desarguesschen Satzes aus dem Pascalschen*, Math. Ann. vol. 61 (1905).

D. Hilbert

1. *Grundlagen der Geometrie*, Teubner, 7th edition, 1930.

P. Jordan, J. von Neumann, and E. Wigner

1. *On an algebraic generalization of the quantum mechanical formalism*, Ann. of Math. (2) vol. 35 (1934) pp. 29–64.

T. P. Kirkman

1. *Schoolgirl problem*, The Lady's and Gentleman's Diary 1850.

A. Kurosch

1. *Ringtheoretische Probleme, die mit dem Burnsideschen Problem über periodische Gruppen in Zusammenhang stehen*, Bull. Acad. Sci. URSS. Sér. Math. vol. 5 (1941) pp. 233–240.

F. Levi

1. *Geometrische Konfigurationen*, Hirzel, Leipzig, 1929.

D. E. Littlewood

1. *Identical relations satisfied in an algebra*, Proc. London Math. Soc. (2) vol. 32 (1931) pp. 312–320.

K. Menger

1. *Bemerkungen zu Grundlagenfragen.* IV, Jber. Deutschen Math. Verein vol. 37 (1928) pp. 309–313.
2. *New foundations of projective and affine geometry*, Ann. of Math. (2) vol. 37 (1936) pp. 456–482.

Ruth Moufang

1. *Zur Struktur der projektiven Geometrie der Ebene*, Math. Ann. vol. 105 (1931) pp. 536–601.
2. *Die Einfuhrung der idealen Elemente in die ebene Geometrie mit Hilfe des Satzes vom vollstandigen Vierseit*, Math. Ann. vol. 105 (1931) pp. 759–778.
3. *Die Schnittpunktsatze des projektiven speziellen Funfecksnetzes in ihrer Abhangigkeit voneinander. (Das A-Netz)*, Math. Ann. vol. 106 (1932) pp. 755–795.

4. *Ein Satz über die Schnittpunktsatze des allgemeinen Funfecksnetzes* [*Das* (*A*, *B*)-*Netz*], Math. Ann. vol. 107 (1932) pp. 124–139.
5. *Die Desarguesschen Satz vom Rang* 10, Math. Ann. vol. 108 (1933) pp. 296–310.
6. *Alternativkörper und der Satz vom vollstandigen Vierseit* (D_9), Abh. Math. Sem. Hamburgischen Univ. vol. 9 (1933) pp. 207–222.
7. *Zur Struktur von Alternativkörpern*, Math. Ann. vol. 110 (1934) pp. 416–430.

B. H. NEUMANN
1. *Identical relations in groups*. I, Math. Ann. vol. 114 (1937) pp. 506–525.
2. *On the commutativity of addition*, J. London Math. Soc. vol. 15 (1940) pp. 203–208.

A. R. RICHARDSON
1. *Equations over a division algebra*, Messenger of Mathematics vol. 57 (1928) pp. 1–6.

A. SPEISER
1. *Theorie der Gruppen von endlicher Ordnung*, 2nd edition (1927), J. Springer, Berlin.

K. G. C. VON STAUDT
1. *Beitrage zur Geometrie der Lage*, vol. 2, 1857.

G. THOMSEN
1. *Grundlagen der Elementargeometrie*, Leipzig, 1933.

O. VEBLEN AND J. H. M. WEDDERBURN
1. *Non-Desarguesian and non-Pascalian geometries*, Trans. Amer. Math. Soc. vol. 8 (1907) pp. 379–388.

O. VEBLEN AND J. W. YOUNG
1. *Projective geometry*, vol. I, Ginn and Co., 1910.

W. WAGNER
1. *Über die Grundlagen der projektiven Geometrie und allgemeine Zahlensysteme*, Math. Ann. vol. 113 (1936–1937) pp. 528–567.

M. ZORN
1. *Theorie der alternativen Ringe*, Abh. Math. Sem. Hamburgischen Univ. vol. 8 (1930) pp. 123–147.
2. *Alternativkörper und quadratische Systeme*, Abh. Math. Sem. Hamburgischen Univ. vol. 9 (1933) pp. 395–402.

YALE UNIVERSITY,
NEW HAVEN, CONN.

ON THE OSCILLATION OF DIFFERENTIAL TRANSFORMS. III
OSCILLATIONS OF THE DERIVATIVE OF A FUNCTION

BY

A. C. SCHAEFFER

1. It has been shown by S. Bernstein that if a function which is defined in an open interval is infinitely often differentiable (belongs to C^∞) and each of its derivatives is of constant sign then the function is analytic in the interval(1). As a possible generalization of this result Pólya raised the following question(2): *If in an open interval a function is infinitely often differentiable and no derivative changes sign more than a given number of times is the function analytic in the interval?* The present paper contains an answer to this question and to several related questions. These results are consequences of an inequality (Theorem I of the present paper) which relates the magnitude of the first derivative of a function to the number of times some higher derivative changes sign. All functions considered in this paper will be supposed real for real values of the variable.

In the case in which $f(x)$ is infinitely often differentiable over the interval $(-\infty, \infty)$ let N_k be the maximum number of sign changes of $f^{(k)}(x)$ over any interval of length α, where α is some arbitrary but fixed positive number. Pólya and Wiener(3) made the additional hypothesis that $f(x)$ is periodic and found that if $N_k = O(1)$ then $f(x)$ is a trigonometric polynomial, while if $N_k = o(k^{1/2})$ then $f(x)$ is an entire function. Szegö obtained a new proof of these results and indeed obtained the sharper theorem(4) that if $N_k < k/\log k$ for large k then $f(x)$ is an entire function. Hille(5) has obtained analogous theorems for a very general class of differential operators and functions which satisfy appropriate conditions.

Let us consider for the moment functions which are bounded and infinitely often differentiable over $(-\infty, \infty)$, but are otherwise completely general. It is shown in the present paper that if $N_k = O(1)$ then $f(x)$ is an entire function of exponential type, and if $N_k = O(\log^\gamma k)$, $\gamma < 1$, then $f(x)$ is an entire function(6). In the case in which N_k is bounded this is a generalization of

Presented to the Society, November 28, 1942; received by the editors February 6, 1943.

(1) S. Bernstein, *Leçons sur les propriétés extrémales* · · ·, Paris, 1926, pp. 196–197.

(2) At the Stanford University Symposium, August 12, 1941.

(3) G. Pólya and N. Wiener, *On the oscillation of the derivatives of a periodic function*, Trans. Amer. Math. Soc. vol. 52 (1942) pp. 249–256.

(4) G. Szegö, *On the oscillation of differential transforms.* I, Ibid. pp. 450–462.

(5) E. Hille, *On the oscillation of differential transforms.* II. *Characteristic series of boundary value problems*, Ibid. pp. 463–497.

(6) A. C. Schaeffer, *Oscillations of the derivatives of a function*, Proc. Nat. Acad. Sci. U. S. A. vol. 28 (1942) pp. 62–63. The results are there stated without proof.

one of the theorems of Pólya and Wiener mentioned above, since an entire function of exponential type which is periodic is a trigonometric polynomial. In the case in which N_k can become infinite as k increases, the results of the present paper overlap those of Pólya and Wiener: we consider a more general class of functions, but, as indicated above, we allow only a much slower rate of growth of N_k.

2. The proofs are based on the following fundamental result.

THEOREM I. *In an interval $a-L<x<a+L$ let $f(x)\in C^n$, $n\geqq 2$, and let $|f(x)|\leqq M$. If*

$$|f'(a)|\geqq (10n)^{2n}M/L \tag{1}$$

then $f^{(n)}(x)$ changes sign at least $n-1$ times in the interval.

Several inequalities are known for the derivatives of functions which satisfy the conditions of Theorem I and have non-negative nth derivatives(7). We show by an example that there is no constant $A=A(n)$ such that the inequality $|f'(a)|\geqq AM/L$ will imply that $f^{(n)}(x)$ has n (or more) variations in sign. For, with $0<\alpha<1$, let

$$f'(x)=\begin{cases}(1/\alpha)(1-x^2/\alpha^2)^n, & |x|\leqq\alpha,\\ 0, & |x|>\alpha.\end{cases}$$

Then if we set $f(0)=0$ it follows from simple calculations that, over the entire real axis, $f(x)\in C^n$, $|f(x)|\leqq 1$, and $f^{(n)}(x)$ has $n-1$ changes in sign. But $f'(0)=1/\alpha$, which is unbounded as α approaches zero.

The following lemma will be sufficient for our purpose, although a more precise formulation is known(8).

LEMMA I (S. Bernstein). *If in a closed interval of length $2L$, $f(x)\in C^n$ and $|f(x)|\leqq 1$ then there is at least one point in the interior of the interval for which*

$$|f^{(n)}(x)|\leqq (2/L)^n n!.$$

One proof, which is undoubtedly known to many, is as follows. Let the interval be $[1, 1]$ and let

$$g(x)=f(x)-\sum_{\nu=1}^{n+1}f(x_\nu)T_{n+1}(x)/((x-x_\nu)T'_{n+1}(x_\nu))$$

where $T_{n+1}(x)=\cos((n+1)\arccos x)=2^n(x-x_1)\cdots(x-x_{n+1})$,

(7) E. Landau, *Über einen Satz von Herrn Esclangon*, Math. Ann. vol. 102 (1930) pp. 177–188. R. P. Boas, *Functions with positive derivatives*, Duke Math. J. vol. 8 (1941) pp. 163–172. R. P. Boas and G. Pólya, *Influence of the signs of the derivatives of a function on its analytic character*, Duke Math. J. vol. 9 (1942) pp. 406–424.

(8) S. Bernstein, loc. cit. p. 10. J. Shohat, *A simple proof of a formula of Tchebycheff*, Tôhoku Math. J. vol. 36 (1932–1933) pp. 230–235. R. P. Boas and G. Pólya, loc. cit. pp. 413–414. With more detail the proof in the text can be made to yield more information.

$x_\nu = \cos((\nu - 1/2)\pi/(n+1))$, is the $(n+1)$st Tchebycheff polynomial. Then $g(x)$ vanishes at $n+1$ or more points, so repeated use of Rolle's Theorem shows that its nth derivative has at least one zero, $g^{(n)}(\alpha) = 0$. Since $T_{n+1}(x)/(x - x_\nu)$ is a polynomial of degree n with leading coefficient 2^n, we obtain

$$0 = f^{(n)}(\alpha) - 2^n n! \sum_{\nu=1}^{n+1} f(x_\nu)/T'_{n+1}(x_\nu).$$

The summation is bounded by 1 since $|T'_{n+1}(x_\nu)| = (n+1)(1-x_\nu^2)^{-1/2} \geqq n+1$, and the lemma follows.

Proof of Theorem I. It is sufficient to consider the case in which $M = 1$ and the interval is $(-1, 1)$. We suppose that $f'(0) \leqq -(10n)^{2n}$ and prove that $f^{(n)}(x)$ must change sign at least $n-1$ times in $(-1, 1)$. According to Lemma I, each closed interval of length $1/(5n)$ lying in $(-1, 1)$ must contain at least one point in its interior where

$$|f^{(k)}(x)| \leqq (20n)^k k! < (20n)^n n^n; \qquad k = 0, 1, 2, \cdots, n. \tag{2}$$

The point x in this inequality depends on k and on the interval.

Let

$$\phi(x) = f(x)(20n)^{-n} n^{-n}.$$

Then

$$\phi'(0) \leqq -10^{2n} 20^{-n} = -5^n \tag{3}$$

and from inequality (2) it follows that each closed interval of length $1/(5n)$ contains at least one point in its interior where

$$|\phi^{(k)}(x)| < 1; \qquad k = 0, 1, 2, \cdots, n. \tag{4}$$

It is to be shown that for $k = 1, 2, \cdots, n$ there are points $x_0, x_1, \cdots, x_{k+1}$, depending on k, such that

$$\begin{gathered} -k/(5n) < x_0 < x_1 < \cdots < x_{k+1} < k/(5n), \\ (-1)^\nu \phi^{(k)}(x_\nu) > 3, \qquad \nu = 1, 2, \cdots, k, \\ |\phi^{(k)}(x_0)| < 1, \qquad |\phi^{(k)}(x_{k+1})| < 1. \end{gathered} \tag{5}$$

In the case $k = n$ these relations will imply that $f^{(n)}(x)$ changes sign at least $n-1$ times, so the theorem will follow. In the case $k = 1$ these relations are a consequence of inequalities (3) and (4) where we have $x_1 = 0$. We now suppose that (5) is true for some k, $1 \leqq k \leqq n-1$, and proceed by induction.

Relations (5) imply that $0 < x_{\nu+1} - x_\nu < 2k/(5n) < 2/5$. The mean value theorem shows that there is a point y_1 such that

$$\phi^{(k)}(x_1) - \phi^{(k)}(x_0) = (x_1 - x_0)\phi^{(k+1)}(y_1), \qquad x_0 < y_1 < x_1.$$

The left-hand side is less than $-3+1 = -2$ so

$$\phi^{(k+1)}(y_1) < -2/(2/5) < -3.$$

Again using the mean value theorem we have (if $k>1$)

$$6 < \phi^{(k)}(x_2) - \phi^{(k)}(x_1) = (x_2 - x_1)\phi^{(k+1)}(y_2), \qquad x_1 < y_2 < x_2,$$

from which we conclude that $\phi^{(k+1)}(y_2)>3$, and so on. Thus there are points

$$-k/(5n) < y_1 < y_2 < \cdots < y_{k+1} < k/(5n)$$

such that

$$(-1)^\nu \phi^{(k+1)}(y_\nu) > 3; \qquad \nu = 1, 2, \cdots, k+1.$$

Then each of the intervals $[y_1-1/(5n), y_1]$ and $[y_{k+1}, y_{k+1}+1/(5n)]$ contains at least one interior point where $|\phi^{(k+1)}|<1$. We call these points y_0 and y_{k+2} respectively and (5) is true with k replaced by $k+1$, which completes the induction.

3. THEOREM II. *If $f(x)\in C^\infty$ in an open interval and there exists an integer p such that no derivative of $f(x)$ changes sign more than p times then the function is analytic in the interval.*

Proof. It is sufficient to show that the function is analytic in the interior of every closed interval in the given interval. (In this way we avoid considering the possibility of the function becoming infinite near the end points.) Also, by a simple transformation the problem is reduced to the case in which the interval is $(-1, 1)$ and $|f(x)| \leqq 1$, $-1<x<1$.

We show by induction that

$$|f^{(k)}(x)| \leqq a^k k!(1-|x|)^{-k}, \qquad -1 < x < 1, \tag{6}$$

for $k=0, 1, 2, \cdots$ where $a=3\{10(p+2)\}^{2(p+2)}$. This inequality is true for $k=0$. If it is true for some $k\geqq 0$ and if $|x_0|<1$, then in the interval $|x-x_0|<(1-|x_0|)/(k+1)$ we have

$$(1-|x|)^{-k} < (1-|x_0|)^{-k}(1-1/(k+1))^{-k} < 3(1-|x_0|)^{-k},$$

so, by (6),

$$|f^{(k)}(x)| < 3a^k k!(1-|x_0|)^{-k}.$$

Theorem I with $L=(1-|x_0|)/(k+1)$ and $n=p+2$ implies that

$$|f^{(k+1)}(x_0)| < \{10(p+2)\}^{2(p+2)}3a^k k!(1-|x_0|)^{-k}(k+1)/(1-|x_0|)$$

or

$$|f^{(k+1)}(x_0)| < a^{k+1}(k+1)!(1-|x_0|)^{-k-1}.$$

Thus (6) is true for all k; and Taylor's expansion of $f(x)$ about any point in the interior of $(-1, 1)$ converges to the function in an entire neighborhood of the point.

THEOREM III. *Let $f(x)\in C^{\infty}(-\infty, \infty)$ and let the number of variations in sign of $f^{(k)}(x)$ in every interval of length α be less than β. If*

$$\overline{\lim_{x\to\pm\infty}} \; |x|^{-1} \log |f(x)| < \infty \tag{7}$$

then $f(x)$ is an entire function of exponential type. Here α and β are arbitrary but fixed positive numbers.

Proof. From (7) we conclude that there are constants A and B such that $|f(x)| < Ae^{B|x|}$ for all real x. There is an integer p such that no derivative of $f(x)$ changes sign more than p times in any interval of length 2.

We suppose that for some k

$$|f^{(k)}(x)| < A\lambda^k e^{B|x|}, \qquad -\infty < x < \infty, \tag{8}$$

where

$$\lambda = e^{B}\{10(p+2)\}^{2(p+2)}, \tag{9}$$

and use induction. Now in an interval of length 2 with center at x_0, inequality (8) shows that

$$|f^{(k)}(x)| < Ae^{B}\lambda^k e^{B|x_0|}.$$

Then Theorem I shows immediately that (8) is true for the next higher integer. Since (8) is true for $k=0$ it is true for all k.

Taylor's expansion of $f(x)$ about the origin then shows that it is an entire function of exponential type λ,

$$|f(z)| \leqq A\sum_{0}^{\infty} \lambda^k |z|^k/k! = Ae^{\lambda|z|},$$

which proves the theorem.

4. The constant λ determines the rate of growth of $f(z)$, and in relation (9) we have an explicit expression for λ in terms of the frequency of the variation in sign of $f^{(k)}(x)$. This expression is not the "best possible," but, at least in the case in which $f(z)$ is bounded on the real axis, a "best possible" inequality for the rate of growth of $f(z)$ can be obtained by use of function theory.

THEOREM IV. *Let $f(x)\in C^{\infty}(-\infty, \infty)$ and let it be bounded. If no derivative of $f(x)$ changes sign more than p times in any interval of length π then $f(x)$ is an entire function of exponential type p,*

$$|f(z)| \leqq Me^{p|y|}.$$

We prove the theorem with p an integer, it will then follow for non-integer p. The proof will depend on two lemmas.

LEMMA II (Pólya-Szegö)(9). *Let $f(z)$ be an entire function which satisfies*

$$f(z) = O(e^{\alpha|z|}), \qquad |f(x)| \leqq M,$$

then

$$|f(z)| \leqq Me^{\alpha|z|}.$$

LEMMA III (S. Bernstein)(10). *Under the conditions of the previous lemma*

$$|f'(x)| \leqq M\alpha, \qquad -\infty < x < \infty,$$

and the equality holds only if

$$f(z) \equiv \sin(\alpha z + \beta).$$

Note that if $f(z)$ satisfies the conditions of Lemma II then its derivative is also an entire function of exponential type α. Hence if to $f(z)$ we apply Lemma III and then Lemma II we obtain $|f'(z)| \leqq M\alpha e^{\alpha|z|}$ which is apparently a stronger statement than Lemma III.

Proof of Theorem IV. A function which satisfies the conditions of Theorem IV is, according to Theorem III, an entire function of exponential type. Let it be of precise type α, that is let

$$\alpha = \overline{\lim_{r\to\infty}}\, r^{-1} \max_{\theta} \log |f(re^{i\theta})|.$$

Then $f(z) = O(e^{(\alpha+\epsilon)|z|})$ if $\epsilon > 0$, and so by Lemma II $|f(z)| \leqq Me^{(\alpha+\epsilon)|z|}$, where M is the upper bound of $f(x)$. Then letting ϵ approach zero, we find that $|f(z)| \leqq Me^{\alpha|z|}$. Let

$$M_k = \sup_{-\infty<x<\infty} |f^{(k)}(x)|$$

and

$$r = \overline{\lim_{k\to\infty}}\, M_{k+1}/M_k.$$

We first show that

$$r = \alpha. \tag{10}$$

Now $f^{(k)}(z)$ is an entire function of exponential type since $f(z)$ is, and $f^{(k)}(z)$ is bounded by M_k on the real axis. Then, from Lemma III we find that $M_{k+1} \leqq \alpha M_k$ and so $r \leqq \alpha$. On the other hand, if $r < \alpha$ there is a constant ρ

(9) G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925, vol. 2, pp. 35–36 and 218–219 (Problems 201, 202). R. Duffin and A. C. Schaeffer, *Some properties of functions of exponential type*, Bull. Amer. Math. Soc. vol. 44 (1938) pp. 236–240, where a more precise formulation is given (for functions which are real on the real axis).

(10) S. Bernstein, loc. cit. pp. 97–102. G. Pólya and G. Szegö, loc. cit. vol. 2 pp. 35 and 218–219 (Problem 201). R. Duffin and A. C. Schaeffer, loc. cit. p. 239. In the discussion of the case of equality it is essential that the functions be real on the real axis.

such that $M_{k+1}/M_k<\rho<\alpha$, $k\geqq k_0$. This clearly implies that $M_k<A\rho^k$, $k=0, 1, 2, \cdots$, for some constant A. Then using Taylor's expansion of $f(z)$ about the origin, we have

$$|f(z)| \leqq A\sum_0^\infty \rho^k |z|^k/k! = Ae^{\rho|z|},$$

which implies that $f(z)$ is of exponential type $\rho<\alpha$. This proves (10).

Let $k_1<k_2<\cdots$ be a sequence of positive integers such that

$$M_{k_\nu+1}/M_{k_\nu} > \alpha(1-1/\nu).$$

Then at some point $f^{(k_\nu+1)}(x)$ will be near its upper bound. Let

$$|f^{(k_\nu+1)}(x_\nu)| > (1-1/\nu)M_{k_\nu+1},$$

and let

$$\phi_\nu(z) = f^{(k_\nu)}(x_\nu+z)/M_{k_\nu}.$$

Then $\phi_\nu(z)$ is an entire function of exponential type α and it is bounded by 1 on the real axis, so by Lemma II

$$|\phi_\nu(z)| \leqq e^{\alpha|y|}. \tag{11}$$

Also, according to the construction,

$$|\phi_\nu'(0)| = |f^{(k_\nu+1)}(x_\nu)|/M_{k_\nu} > (1-1/\nu)M_{k_\nu+1}/M_{k_\nu} > \alpha(1-1/\nu)^2.$$

Inequality (11) together with Vitali's convergence theorem implies that there is a subsequence of the functions $\phi_\nu(z)$ which converges to an entire function $\phi(z)$ with the properties that

$$|\phi(z)| \leqq e^{\alpha|y|}$$

and

$$|\phi'(0)| \geqq \alpha.$$

The case of equality in Lemma III then shows that

$$\phi(z) \equiv \pm \sin \alpha z.$$

If $\alpha>p$ then $\phi(x)$ would change sign at least $p+1$ times in an interval $(-\epsilon, \pi-\epsilon)$ where ϵ is some sufficiently small positive number, and then $f^{(k_\nu)}(x)$ would change sign at least $p+1$ times in an interval of length π if k_ν is large. This is impossible, so $\alpha\leqq p$ and the theorem follows.

5. In the following let N_k be such that $f^{(k)}(x)$ does not change sign more than N_k times in any interval of length α, where α is an arbitrary but fixed positive number.

THEOREM V. *Let $f(x)\in C^\infty(-\infty, \infty)$ and let $N_k=O(\log^\gamma k)$, $\gamma<1$. If*

$$\overline{\lim_{x\to\pm\infty}} \; |x|^{-1} \log |f(x)| < \infty$$

then $f(x)$ is an entire function of order not exceeding one.

It is sufficient to consider the case in which $\alpha=2$. There are constants M_0 and B such that

$$|f(x)| < M_0 e^{B|x|}, \qquad -\infty < x < \infty.$$

If x_0 is any real number then $f(x)$ is bounded by $M_0 e^B e^{B|x_0|}$ in the interval $|x-x_0|<1$. Then in virtue of the conditions of this theorem there is an integer k such that $N_k<k-2$, and so Theorem I gives a dominant for $f'(x)$ of the form $|f'(x)| < M_1 e^{B|x|}$, $-\infty<x<\infty$. Repetition of this argument shows that for $k=0, 1, 2, \cdots$

$$|f^{(k)}(x)| < M_k e^{B|x|}, \qquad -\infty < x < \infty, \tag{12}$$

and the object is to prove that the M_k can be chosen to increase at a sufficiently slow rate so that $f(x)$ is an entire function of order no larger than one.

If $\gamma<\mu<1$ then for large k,

$$N_k < \log^\mu k.$$

Also, if $\mu<\lambda<1$ and we define n by

$$n = [\log^\lambda k] \tag{13}$$

then for large k,

$$N_{k+n} < \log^\mu (k + \log^\lambda k) < n - 2.$$

If k is large (12) shows that $f^{(k)}(x)$ is bounded by $M_k e^B e^{B|x_0|}$ in the interval $|x-x_0|<1$, so, applying Theorem I to the function $f^{(k)}(x)$ with n defined by (13), we have

$$|f^{(k+1)}(x_0)| < (10n)^{2n} M_k e^B e^{B|x_0|}.$$

Thus

$$M_{k+1} \leqq e^B (10n)^{2n} M_k,$$

and after some simplification this becomes

$$M_{k+1} \leqq k^\epsilon M_k, \qquad \epsilon > 0,$$

for large k. Finally,

$$M_k = O((k!)^\epsilon)$$

for every $\epsilon>0$, and this shows that $f(x)$ is an entire function of order no greater than one.

STANFORD UNIVERSITY,
STANFORD UNIVERSITY, CALIF.

ON HOMOLOGICAL SITUATION PROPERTIES OF COMPLEXES AND CLOSED SETS

BY
PAUL ALEXANDROFF

The purpose of this paper is to find and to study topological invariants which connect the homological properties of a space K with those of its closed subset A and of the open complement $G=K\setminus A$, and thus help to characterize from the homological point of view the *situation of A in K*.

In the case when K is simply connected (that is when the Betti groups of K are zero) the problem is solved by the duality theorems of Alexander, Pontrjagin, and Kolmogoroff, which determine the Betti groups of G through the Betti groups of A. In the other special case when K is a manifold the first duality theorems have been obtained by Pontrjagin [11](1) and Lefschetz [9] in 1927–1928. All these results are special cases of the general theory to which the present paper is devoted and which gives the very elementary construction (of the so-called extension- and intersection-homomorphisms, section 11) dominating the whole variety of duality and other situation properties.

The complete combinatorial theory is given in Chapter I for an arbitrary cell complex K and its closed subcomplex A. In Chapter III the same theory is generalized for locally bicompact normal spaces K and their closed sets A; this generalization is based on an approximation process developed in Chapter II. Chapter IV deals with manifolds and gives an elementary proof (combinatorial in character) of the Alexander-Pontrjagin duality in its most general form.

All main results obtained are completely formulated in the first four sections of Chapter I (sections 11–14). Numerical consequences are given in section 16. Section 18 deals with the Phragmén-Brouwer problem, while in section 19 some quite elementary examples are given as illustration.

The elementary known facts and notations used throughout this paper are systematized in the Introduction; its first part contains the group theoretical material, the second the needed information on complexes. Thus the present paper is practically independent of the previous literature on related subjects. There are only few references to my paper [1], and each of them may be read without reading the rest of that paper.

INTRODUCTION

§1. GROUPS AND THEIR CHARACTER GROUPS

1. Groups, homomorphisms, duality. All groups considered are commuta-

Presented to the Society, April 24, 1943; received by the editors February 17, 1943.
(1) Numbers in brackets refer to the bibliography at the end of the paper.

tive and written additively. In topological groups only closed subgroups and continuous homomorphisms are allowed. The difference- (or factor-) group of a group X over its subgroup X_0 is denoted by $X-X_0$.

Let σ be a homomorphism of the group X into the group Y; let $Y_1=\sigma X$ be the *image* of X under σ and $X_0=\sigma^{-1}0$ (or $\sigma^{-1}0_Y$) be the *kernel* of the homomorphism σ, that is, the subgroup of all elements of X mapped by σ on the zero element 0_Y of Y. Then we often use the picture

$$\begin{array}{c} X \supseteq X_0 \\ \downarrow \\ Y \supseteq Y_1 \end{array}.$$

We denote by I the additive group of all integers, by $\bar{I}$ the additive continuous (bicompact) group of all real numbers reduced modulo 1.

If B is the character group of the group A in the sense of Pontrjagin [10] then A is the character group of B, and we write

$$A \mid B. \tag{1.1}$$

The groups A and B are called in this case *dual* and the relation (1.1) is called *a duality*.

2. **Scalar products.** If $A \mid B$, $a \in A$, $b \in B$, then the character b of the group A and the character a of the group B map respectively the elements $a \in A$ and $b \in B$ on the same element

$$ba = ab \in \bar{I}$$

called the *product* of a and b.

Other cases in which products are defined are:

1. When a and b are elements of the same (commutative) ring.

2. When a is an element of an arbitrary group A, while b is an integer. Then for $b>0$

$$ab = ba = a + a + \cdots + a \qquad (b \text{ times}),$$

$$a(-b) = (-b)a = -(ab),$$

and $ab=ba=0$ for $b=0$.

The *scalar product*

$$(x \cdot y) = \sum a_i b_i$$

of two linear forms $x=\sum a_i t_i$ and $y=\sum b_i t_i$ is defined if the products $a_i b_i$ are defined.

3. **Annihilators.** Let $X \mid Y$; let $A \subseteq X$ and $B \subseteq Y$ be subgroups of X and Y respectively. If B is the *annihilator* of A (in Y), that is, the subgroup of all elements y of Y satisfying the relation $xy=0$ whenever $x \in A$, then A is the annihilator of B in X and we write

$$X \supseteq A \perp B \subseteq Y,$$

this relation being called an *annihilation.*

A fundamental theorem of the character theory is:

THEOREM 3. *The annihilation* $X \supseteq A \perp B \subseteq Y$ *implies the dualities*

$$(X - A) \mid B, \qquad (Y - B) \mid A.$$

4. Conjugate homomorphisms. The fundamental lemma. Let

$$X \mid \overline{X}, \qquad Y \mid \overline{Y}.$$

The homomorphisms

$$\sigma \text{ of } X \text{ into } Y; \qquad \bar{\sigma} \text{ of } \overline{Y} \text{ into } \overline{X}$$

are called *conjugate* if $(\sigma x \cdot \bar{y}) = (x \cdot \bar{\sigma}\bar{y})$, whenever $x \in X$, $\bar{y} \in \overline{Y}$. In this case we write

$$\sigma \left| \begin{matrix} X & \overline{X} \\ \downarrow & \uparrow \\ Y & \overline{Y} \end{matrix} \right| \bar{\sigma}$$

or—if the images $Y_1 = \sigma X$, $\overline{X}_1 = \bar{\sigma}\overline{Y}$ and the kernels $X_0 = \sigma^{-1}0_Y$, $\overline{Y}_0 = \bar{\sigma}^{-1}0_{\overline{X}}$ are given—

$$(4.1) \qquad \sigma \left| \begin{matrix} X \supseteq X_0 & \overline{X}_1 \subseteq \overline{X} \\ \downarrow \quad \nearrow & \uparrow \quad \nwarrow \\ Y \supseteq Y_1 & \overline{Y}_0 \subseteq \overline{Y} \end{matrix} \right| \bar{\sigma}.$$

Most of this paper is based on the

FUNDAMENTAL LEMMA 4. *From* (4.1) *follows*

$$(4.2) \qquad \begin{cases} X \supseteq X_0 & \perp \quad \overline{X}_1 \subseteq \overline{X} \\ \downarrow \quad \nearrow & \qquad \uparrow \quad \nwarrow \\ Y \supseteq Y_1 & \perp \quad \overline{Y}_0 \subseteq \overline{Y} \end{cases}$$

(*where the diagonal line means the duality* $\overline{X}_1 \mid Y_1$) *and*

$$(4.3) \qquad \begin{matrix} X_0 \mid (\overline{X} - \overline{X}_1) \\ (Y - Y_1) \mid \overline{Y}_0. \end{matrix}$$

The dualities (4.3) follow from Theorem 3 and the annihilations in (4.2), while the duality $\overline{X}_1 \mid Y_1$ follows from Theorem 3 by means of the isomorphism $Y_1 \approx X - X_0$. Thus we need to prove only the annihilations in (4.2). It is sufficient to prove the first annihilation

$$X \supseteq X_0 \perp \overline{X}_1 \subseteq \overline{X}.$$

To this end let $x \in X_0$, $\bar{y} \in \overline{Y}$. Then

$$(x \cdot \bar{\sigma}\bar{y}) = (\sigma x \cdot \bar{y}) = 0.$$

On the other hand, if $\sigma x \neq 0$, then there exists a $\bar{y} \in \overline{Y}$ with

$$(x \cdot \bar{\sigma}\bar{y}) = (\sigma x \cdot \bar{y}) \neq 0.$$

Lemma 4 is proved.

5. **Remark concerning the rational case.** Let $\mathfrak{R}$ be the field of all rational numbers, $\mathfrak{R}^n$ the corresponding n-dimensional linear space, that is, the group of all linear forms with n variables and rational coefficients; the group $\mathfrak{R}^n$ is considered as a group with operators, the latter being rational numbers. Thus only linear subspaces are allowed as subgroups. If we understand now by a character of $\mathfrak{R}^n$ a homomorphism of $\mathfrak{R}^n$ into $\mathfrak{R}$, the "character group" of $\mathfrak{R}^n$ is again $\mathfrak{R}^n$, as follows from the definition $y(x) = x(y) = (x \cdot y)$ for any $x \in \mathfrak{R}^n$, $y \in \mathfrak{R}^n$, the scalar product being understood in the most elementary sense. Thus the duality relation | turns in this case into the identity and Lemma 4 still holds but now becomes trivial.

§2. Concerning cell complexes

6. **Cell spaces and their subspaces.** A cell space (Tucker [14], Kolmogoroff [8]) is a finite set of elements, called *cells*, satisfying the following conditions:

1. To any cell corresponds a non-negative integer, the *dimension* of the cell (denoted by a superscript: t^r is an r-dimensional cell or an *r-cell*).

2. To any cell t^r corresponds a well defined cell $-t^r$ of the same dimension r, and we have $-(-t^r) = t^r$; the cells t^r and $-t^r$ are called *opposite* (to each other).

3. To any two cells t^r and t^{r-1} of dimensions r and $r-1$, respectively, corresponds an integer $(t^r : t^{r-1})$, the *incidence coefficient* of t^r and t^{r-1}, under the conditions:

$$(-t^r : t^{r-1}) = (t^r : -t^{r-1}) = -(t^r : t^{r-1}).$$

It is convenient to suppose that in any pair of opposite cells of the cell space K a *definite* cell is denoted by t_i^r and the other by $-t_i^r$. Then we put

$$\epsilon_{ij}^r = (t_i^r : t_j^{r-1}).$$

Let $t_1^r \in K$ and $t_2^s \in K$ be two cells, and $s < r$. We write $t_1^r > t_2^s$ or $t_2^s < t_1^r$ if cells $t_{j_1}^{r-1} \in K, \cdots, t_{j_{r-s-1}}^{s+1} \in K$ can be found so that

$$(t_i^r : t_{j_1}^{r-1}) \neq 0,\ (t_{j_1}^{r-1} : t_{j_2}^{r-2}) \neq 0, \cdots,\ (t_{j_{r-s-1}}^{s+1} : t_k^s) \neq 0.$$

If $t_2^s < t_1^r$ we say that t_2^s is a *face* of t_1^r.

A cell space Q is called a *subspace* of the cell space K if Q is a subset of K and

(a) two cells are opposite in Q if and only if they are opposite in K;

(b) the dimension of a cell and the incidence coefficient of two cells in Q are the same as in K.

A subspace Q of the cell space K is called: *closed*, if $t\in Q$, $t'<t$ (in K) implies $t'\in Q$; *open*, if $t\in Q$, $t'>t$ (in K) implies $t'\in Q$.

7. **Chains and boundaries.** An *r-dimensional chain* (or an *r-chain*) on a cell space K with respect to the group X (the "*coefficient-group*") is a function x^r with values in X defined on the set of all r-dimensional cells of K, taking opposite values on opposite cells. The additive group of all r-dimensional chains on K with respect to X is denoted by $L^r_K(X)$ or by L^r_K. Chains with respect to the group I of all integers are called *integer chains*.

The chain taking the value $a\in X$ on the cell $t^r\in K$ and the value zero on all cells different from $\pm t^r$ is denoted by at^r (in particular, the integer chain taking the value 1 on t^r, the value -1 on $-t^r$ and the value 0 elsewhere is identified with the cell t^r). This convention allows one to set for any chain $x^r\in L^r_K$

$$x^r = \sum a_i t^r_i$$

where a_i is the value taken by x^r on t^r_i.

The *scalar product*

$$(x^r \cdot y^r) = \sum a_i b_i$$

is defined for two chains $x^r=\sum a_i t^r_i$ and $y^r=\sum b_i t^r_i$ under the hypotheses for coefficient-groups mentioned in section 2. In particular the scalar product $(x^r \cdot t^r_i)$ is always defined and equal to the value of the chain x^r on the cell t^r_i.

Let x^r be any chain of the cell space K. The $(r-1)$-chain taking the value $\sum_i \epsilon^r_{ij}(x^r \cdot t^r_i)$ on any $t^{r-1}_j\in K$ is denoted by Δx^r and called the Δ-boundary (or the lower boundary) of x^r (on K). The $(r+1)$-chain taking the value $\sum_i \epsilon^{r+1}_{ki}(x^r \cdot t^r_i)$ on $t^{r+1}_k\in K$ is denoted by ∇x^r and called the ∇- (or upper) boundary of x^r on K; sometimes we write $\Delta_K x^r$, $\nabla_K x^r$ instead of Δx^r, ∇x^r to show the cell space on which Δ, ∇ operate.

Obviously

$$(\nabla t^r_i \cdot t^{r+1}_\lambda) = (\Delta t^{r+1}_\lambda \cdot t^r_i) = \epsilon^{r+1}_{\lambda i},$$

and thus for any two chains x^r and y^{r+1}

$$(\nabla x^r \cdot y^{r+1}) = (x^r \cdot \Delta y^{r+1}) \tag{7.1}$$

provided that these scalar products have a sense (section 2).

Let $X \mid \overline{X}$, $L^r_K = L^r_K(X)$, $\overline{L}^r_K = L^r_K(\overline{X})$. Denote by

$$H^r_K = \nabla L^{r-1}_K; \qquad \overline{H}^r_K = \Delta \overline{L}^{r+1}_K$$

the images; by Z^r_K, $\overline{Z}^r_K$ the kernels of the homomorphisms

$$\nabla(\text{of } L^{r-1}_K \text{ into } L^r_K); \qquad \Delta(\text{of } \overline{L}^{r+1}_K \text{ into } \overline{L}^r_K).$$

The elements of Z^r_K (of $\overline{Z}^r_K$) are called r-dimensional ∇-cycles (Δ-cycles); among all elements x^r of L^r_K (of $\overline{L}^r_K$) they are characterized by the condition

$$\nabla x^r = 0; \qquad \Delta x^r = 0.$$

From (7.1) and the Fundamental Lemma 4 follows

(7.2)
$$\nabla \left| \begin{array}{ccccc} L^{r-1}_K & \supseteq & Z^{r-1}_K & \perp & H^{r-1}_K \subseteq \overline{L}^{r-1}_K \\ L^r_K & \supseteq & H^r_K & \perp & \overline{Z}^r_K \subseteq \overline{L}^r_K \end{array} \right| \Delta,$$

(7.3)
$$(L^r_K - H^r_K) \,|\, \overline{Z}^r_K.$$

8. **Cell complexes.** A cell space is called a cell *complex* if for each of its cells $\Delta\Delta t^r = 0$, that is $\sum_j \epsilon^r_{ij}\epsilon^{r-1}_{jk} = 0$ for every r, i, k. On a cell complex $\Delta\Delta x^r = 0$, $\nabla\nabla x^r = 0$ for every chain x^r and therefore $H^r_K \subseteq Z^r_K$, $\overline{H}^r_K \subseteq \overline{Z}^r_K$. The elements of H^r_K, $\overline{H}^r_K$ are called *bounding cycles* (∇- and Δ-, respectively).

9. **Betti groups.** These are the groups

$$\nabla^r_K = \nabla^r_K(X) = Z^r_K - H^r_K; \qquad \Delta^r_K = \Delta^r_K(\overline{X}) = \overline{Z}^r_K - \overline{H}^r_K;$$

the first is called the r-dimensional ∇- (or upper) group and the second is called the r-dimensional Δ- (or lower) group of K with respect to the given coefficient-group (X for ∇ and $\overline{X}$ for Δ). The elements of ∇^r_K (of Δ^r_K) are called r-dimensional ∇-classes (Δ-classes) of K. Two ∇-cycles (Δ-cycles) z^r_1 and z^r_2 are called *homologous* to each other on K if they belong to the same ∇-class (Δ-class); we write in this case $z^r_1 \sim z^r_2$ on K.

Let us prove the duality

(9)
$$\nabla^r_K \,|\, \Delta^r_K.$$

We recall from the character theory that from $A \,|\, \overline{A}$, $C \,|\, \overline{C}$, $C = A - B$ follows $B \,|\, \overline{A} - \overline{C}$. Apply this to the Emmy Noether isomorphism

$$(L^r_K - H^r_K) - (Z^r_K - H^r_K) \approx L^r_K - Z^r_K$$

and take from (7.3), (7.2) (with $r+1$ instead of r)

$$(L^r_K - H^r_K) \,|\, \overline{Z}^r_K, \qquad L^r_K - Z^r_K \approx H^{r+1}_K \,|\, \overline{H}^r_K.$$

We get

$$(Z^r_K - H^r_K) \,|\, (\overline{Z}^r_K - \overline{H}^r_K),$$

q.e.d.

10. **Closed and open subcomplexes.** Let Q be a closed subspace of the cell complex K and $t^r \in Q$. Then $\Delta_Q t^r = \Delta_K t^r$ and $\Delta_Q \Delta_Q t^r = \Delta_K \Delta_K t^r = 0$; thus Q is a cell complex.

Let Q be an open subspace of the cell complex K. Take any chain x^r on Q and define the chain $E^Q_K x^r$, called the *trivial extension* of x^r over K, by setting

$E_K^Q x^r = x^r$ on Q, $E_K^Q x^r = 0$ on $K \setminus Q$. Then since Q is open on K, we have for any $t^{r-1} \in Q$

$$(\Delta_Q x^r \cdot t^{r-1}) = (\Delta_K E_K^Q x^r \cdot t^{r-1}).$$

It follows that

$$(\Delta_Q \Delta_Q x^r \cdot t^{r-2}) = (\Delta_K \Delta_K E_K^Q x^r \cdot t^{r-2}) = 0$$

for any $t^{r-2} \in Q$, and Q is again a cell complex.

Thus *closed as well as open subspaces of a cell complex are cell complexes.*

CHAPTER I. COMPLEXES

§3. THE GENERAL THEORY

11. **The extension- and the intersection-homomorphisms.** Let Q be a subcomplex of the cell complex K and x_Q^r a chain on Q. Any chain on K taking on the cells of Q the same values as the chain x_Q^r is called an *extension* of Q (over K). Among these extensions the *trivial extension* $E_K^Q x_Q^r$ takes the value zero on $K \setminus Q$.

If we assign to any chain on Q its trivial extension $E_K^Q x_Q^r$, we get the *extension-isomorphism* of L_Q^r into L_K^r.

REMARK. If there is no possibility of confusion, we shall identify chains with their trivial extensions and thus consider the extension-isomorphism as the identical isomorphism of L_Q^r into $L_K^r \supseteq L_Q^r$.

Now, let x_K^r be any chain on K. We denote by $J_Q^K x_K^r$ the chain *on* Q taking on the cells of Q the same values as the chain x_K^r. If we assign to any chain x^r on K the chain $J_Q^K x_K^r$, we get a homomorphism J_Q^K of L_K^r into L_Q^r called the *intersection-homomorphism.*

Obviously, if $X \mid \overline{X}$, $L_K^r = L_K^r(X)$, $\overline{L}_K^r = L_K^r(\overline{X})$, then

$$(11.1) \qquad J_Q^K \Big| \begin{matrix} L_K^r & \overline{L}_K^r \\ \downarrow & \uparrow \\ L_Q^r & \overline{L}_Q^r \end{matrix} \Big| E_K^Q.$$

From now to the end of this chapter K is a fixed cell complex, A is a fixed closed subcomplex of K and $G = K \setminus A$ is the complementary open subcomplex.

It is easily seen that for any chains x_A^r on A, x_G^r on G, x_K^r on K:

$$(11.2) \qquad \Delta E_K^A x_A^r = E_K^A \Delta x_A^r; \qquad \nabla E_K^G x_G^r = E_K^G \nabla x_G^r,$$

$$(11.3) \qquad \nabla J_A^K x_K^r = J_A^K \nabla x_K^r; \qquad \Delta J_G^K x_K^r = J_G^K \Delta x_K^r.$$

11.4. Therefore, the extension isomorphism maps identically

Δ-cycles of A	∇-cycles of G
on Δ-cycles of K	on ∇-cycles of K

while bounding cycles are mapped on bounding cycles. Thus the extension-

isomorphism generates a homomorphism called the ***extension-homomorphism***:

$$E_K^A \text{ of } \Delta_A^r \text{ into } \Delta_K^r; \qquad E_K^G \text{ of } \nabla_G^r \text{ into } \nabla_K^r.$$

11.5. The intersection-homomorphism maps ∇-cycles of K on ∇-cycles of A; Δ-cycles of K on Δ-cycles of G (while bounding cycles are mapped on bounding cycles) and thus generates a homomorphism, also called *the intersection-homomorphism*:

$$J_A^K \text{ of } \nabla_K^r \text{ into } \nabla_A^r; \qquad J_G^K \text{ of } \Delta_K^r \text{ into } \Delta_G^r.$$

12. **The groups Δ_{KA}^r, ∇_{KG}^r; $\Delta_{A:K}^r$, $\nabla_{G:K}^r$; ∇_{AK}^r, Δ_{GK}^r.** We write 0_K for the zero element of both Δ_K^r and ∇_K^r, and use 0_A, 0_G in an analogous sense. Now *we define*

$$\Delta_{KA}^r = E_K^A \Delta_A^r \subseteq \Delta_K^r; \qquad \nabla_{KG}^r = E_K^G \nabla_G^r \subseteq \nabla_K^r \qquad \text{(images)}, \tag{12.1}$$

$$\Delta_{A:K}^r = (E_K^A)^{-1} 0_K \subseteq \Delta_A^r; \qquad \nabla_{G:K}^r = (E_K^G)^{-1} 0_K \subseteq \nabla_G^r \qquad \text{(kernels)}, \tag{12.2}$$

$$\nabla_{AK}^r = J_A^K \nabla_K^r \subseteq \nabla_A^r; \qquad \Delta_{GK}^r = J_G^K \Delta_K^r \subseteq \Delta_G^r \qquad \text{(images)}, \tag{12.3}$$

and *prove* for the kernels of the intersection-homomorphism:

$$(J_A^K)^{-1} 0_A = \nabla_{KG}^r; \qquad (J_G^K)^{-1} 0_G = \Delta_{KA}^r. \tag{12.4}$$

We prove only the left-hand formula (writing J for J_A^K and E for E_K^A).

1. If $z^r \in Z_K^r$, $z^r = z_G^r + h^r$ with $z_G^r \in Z_G^r$, $h^r \in H_K^r$, then $Jz^r = Jh^r \in H_A^r$.

2. If $z^r \in Z_K^r$, $Jz^r \in H_A^r$, then $z^r = z_G^r + h^r$ with $z_G^r \in Z_G^r$, $h^r \in H_K^r$.

Let us prove 2. We have $Jz^r = \Delta x_A^{r-1}$, $x_A^{r-1} \in L_A^{r-1}$. For any $t_A^r \in A$

$$((z^r - \nabla E x_A^{r-1}) \cdot t_A^r) = (z^r \cdot t_A^r) - (\nabla E x_A^{r-1} \cdot t_A^r).$$

Now

$$(z^r \cdot t_A^r) = (Jz^r \cdot t_A^r) = (\nabla x_A^{r-1} \cdot t_A^r),$$

$$(\nabla E x_A^{r-1} \cdot t_A^r) = (J \nabla E x_A^{r-1} \cdot t_A^r) = (\nabla J E x_A^{r-1} \cdot t_A^r) = (\nabla x_A^{r-1} \cdot t_A^r).$$

Hence

$$((z^r - \nabla E x_A^{r-1}) \cdot t_A^r) = 0 \qquad \text{for } t_A^r \in A$$

so that the cycle $z^r - \nabla E x_A^{r-1}$ is on G. As $\nabla E x_A^{r-1} \in H_K^r$ the assertion 2 and thus the formula (12.4) are proved.

13. **The geometrical meaning of the groups Δ_{KA}^r, ∇_{KG}^r; $\Delta_{A:K}^r$, $\nabla_{G:K}^r$; ∇_{AK}^r, Δ_{GK}^r.**

13.1. The subgroup

$$\Delta_{KA}^r \text{ of } \Delta_K^r; \qquad \nabla_{KG}^r \text{ of } \nabla_K^r$$

defined as the image of Δ_A^r (of ∇_G^r) under the extension-homomorphism has for its elements those Δ-classes (∇-classes) of K which contain Δ-cycles on A (∇-cycles on G).

13.2. Take all those

$$\Delta\text{-classes } \mathfrak{z}^r_A \in \Delta^r_A; \qquad \nabla\text{-classes } \mathfrak{z}^r_G \in \nabla^r_G$$

whose elements bound on K. The resulting subgroup of Δ^r_A (of ∇^r_G) is the group $\Delta^r_{A:K}$ (the group $\nabla^r_{G:K}$), defined as the kernel of the extension-homomorphism.

DEFINITION 13.3. *A ∇-cycle on A (a Δ-cycle on G) is called extensible (over K) if among its extensions over K there are ∇-cycles (Δ-cycles) on K.*

13.4. Then the group ∇^r_{AK} (the group Δ^r_{GK}) defined as the image of ∇^r_K (of Δ^r_K) under the intersection-homomorphism is the difference-group of the group of all r-dimensional extensible ∇-cycles on A (Δ-cycles on G) over the subgroup of cycles bounding on A (on G).

14. **The results.** From the definition of the groups involved, from (11.1) and (12.4) follows by the Fundamental Lemma 4:

$$(14.1) \qquad \begin{array}{c|c|c} E^A_K & \Delta^r_A \supseteq \Delta^r_{A:K} \perp \nabla^r_{AK} \subseteq \nabla^r_A & J^K_A \\ J^K_G & \Delta^r_K \supseteq \Delta^r_{KA} \perp \nabla^r_{KG} \subseteq \nabla^r_K & E^G_K \\ & \Delta^r_G \supseteq \Delta^r_{GK} \perp \nabla^r_{G:K} \subseteq \nabla^r_G & \end{array}$$

and

$$(14.2) \qquad \Delta^r_{A:K} \mid (\nabla^r_A - \nabla^r_{AK}); \qquad \nabla^r_{G:K} \mid (\Delta^r_G - \Delta^r_{GK}).$$

The picture (14.1) contains

1. two pairs of isomorphisms: the first pair

$$\Delta^r_A - \Delta^r_{A:K} \approx \Delta^r_{KA}; \qquad \nabla^r_G - \nabla^r_{G:K} \approx \nabla^r_{KG};$$

the second pair

$$\Delta^r_K - \Delta^r_{KA} \approx \Delta^r_{GK}; \qquad \nabla^r_K - \nabla^r_{KG} \approx \nabla^r_{AK},$$

2. the pair of dualities

$$\Delta^r_{KA} \mid \nabla^r_{AK}; \qquad \nabla^r_{KG} \mid \Delta^r_{GK};$$

3. three annihilations: the central annihilation

$$\Delta^r_K \supseteq \Delta^r_{KA} \perp \nabla^r_{KG} \subseteq \nabla^r_K$$

and the pair of annihilations

$$\Delta^r_A \supseteq \Delta^r_{A:K} \perp \nabla^r_{AK} \subseteq \nabla^r_A; \qquad \nabla^r_G \supseteq \nabla^r_{G:K} \perp \Delta^r_{GK} \subseteq \Delta^r_G.$$

In each pair of relations (isomorphisms, dualities, annihilations) the one relation is obtained from the other by interchanging Δ and ∇, A and G.

From these relations the first pair of isomorphisms is a triviality, being contained in the very definition of the groups involved.

The dualities in (14.1) and (14.2) are algebraic consequences of the annihilations and the latter follow from the fact that the extension- and intersection-homomorphisms are conjugate.

The central annihilation may be brought obviously into the form of the following

REMOVING THEOREM. *In order that an r-dimensional ∇-cycle (Δ-cycle) on K be homologous on K to a*

∇-*cycle on* G, Δ-*cycle on* A

it is necessary and sufficient that its scalar product with each r-dimensional Δ-cycle on A; r-dimensional ∇-cycle on G be equal to zero(2).

In addition to (14.1) and (14.2) we have

THE MAIN DUALITY THEOREM.

$$\Delta^r_{A:K} \mid \nabla^{r+1}_{G:K}. \tag{14.3}$$

We give to this fundamental duality the form of a pair of isomorphisms by means of the following *definitions*:

$$\nabla^r_{A:K} = \nabla^r_A - \nabla^r_{AK}; \qquad \Delta^r_{G:K} = \Delta^r_G - \Delta^r_{GK}.$$

REMARK. The groups $\nabla^r_{A:K}$, $\Delta^r_{G:K}$ can also be defined as the difference groups of the groups of all r-dimensional ∇-cycles of A (Δ-cycles of G) over the subgroups of all extensible cycles.

With these definitions we deduce from (14.2) the dualities

$$\Delta^r_{A:K} \mid \nabla^r_{A:K}; \quad \nabla^r_{G:K} \mid \Delta^r_{G:K} \tag{14.2'}$$

and transform the duality (14.3) into the third pair of isomorphisms:

$$\Delta^r_{A:K} \approx \Delta^{r+1}_{G:K}; \qquad \nabla^r_{A:K} \approx \nabla^{r+1}_{G:K}. \tag{14.3'}$$

The groups Δ^r_{KA}, ∇^r_{KG}; $\Delta^r_{A:K}$, $\nabla^r_{G:K}$; ∇^r_{AK}, Δ^r_{KG}; $\nabla^r_{A:K}$, $\Delta^r_{G:K}$ are called the *groups of the figure K, A, G.* It follows from (14.1)–(14.3) that *they are all completely determined* if we know for instance the groups Δ^r_K, Δ^r_A, Δ^r_{KA} (for all r).

In Chapters II and III the definitions of the extension- and of the intersection-homomorphisms as well as the definitions of the groups of the figure K, A, G are generalized to the case of an arbitrary locally bicompact normal space K and its subsets: the closed set $A \subseteq K$ and its open complement $G = K \setminus A$. We show in these chapters that *all relations* (14.1)–(14.3') *hold in that general set theoretical case.*

(2) I am indebted to L. Pontrjagin who first formulated the Removing Theorem and many times pointed out to me (as did also L. Lusternik) the desirability of proving it under the most general conditions.

In particular, the duality theorem (14.3) (or the equivalent pair of isomorphisms (14.3′)) is the definitive generalization of the duality theorem of Alexander-Kolmogoroff: If the space K is simply connected in the dimension r (that is, if one and therefore both of the equivalent conditions $\Delta'_K=0$, $\nabla'_K=0$ hold), then obviously $\Delta'_{A:K}=\Delta'_A$, $\nabla'_{G:K}=\nabla'_G$. Hence, *if K is simply connected in the dimensions r and $r+1$ then* (14.3′) *turns into the Alexander-Kolmogoroff isomorphisms*

$$\Delta'_A \approx \Delta'^{+1}_G, \qquad \nabla'_A \approx \nabla'^{+1}_G.$$

Let us return to the combinatorial case, that is, let K be a cell complex, A a closed subcomplex of K. Taking as coefficient-group *the group I_m of all integers reduced modulo m* and denoting by $\Delta'_K(m)$, $\Delta'_A(m)$, $\Delta'_{A:K}(m)$, $\nabla'_{KG}(m)$ and so on the corresponding groups (of K, of A, of G, of K, A, G) we see by the first pair of isomorphisms that the order of the group $\Delta'_A(m)$ is equal to the product of the orders of $\Delta'_{KA}(m)$ and $\Delta'_{A:K}(m)$ and thus is completely determined by the last two groups. Now by a theorem of M. Bockstein [6, p. 373] all Betti groups of a complex K are completely determined by the orders of the groups $\Delta'_K(m)$ taken for all m and all r. Thus the groups Δ'_A are determined by the groups $\Delta'_{KA}(m)$ and $\Delta'_{A:K}(m)$ (taken for all m and all r). An analogous result holds for ∇'_G, while Δ'_K is determined by the $\Delta'_{KA}(m)$ and $\nabla'_{KG}(m)$.

14.4. Hence *if K is a cell complex, A a closed subcomplex of K and $G=K\setminus A$, then the groups Δ'_A and Δ'_G (and all groups of the figure K, A, G) for any r and any coefficient-group are completely determined by the groups $\Delta'_K(m)$, $\Delta'_{KA}(m)$ and $\Delta'_{A:K}(m)$ taken for all m and all r (while the groups Δ'_K are determined by $\Delta'_{KA}(m)$ and $\nabla'_{KG}(m)$ taken also for all m and all r).*

In Chapter III, section 40, we prove that Theorem 14.4 still holds if we understand by K and A topological images of (finite) polyhedrons (in the terminology of [4, p. 128]). Thus *for topological images of finite polyhedrons the homological situation-problem may be considered as being completely solved.*

REMARK. The third pair of isomorphisms is given for the first time in my notes [2].

15. **The operator ∇E^A_K and the proof of the third pair of isomorphisms.** To any ∇-cycle z'_A on A corresponds the ∇-cycle $\nabla E^A_K z'_A$ on G which by its very definition bounds on K. Let us prove: *If z'_A is extensible, then $\nabla E^A_K z'_A$ bounds on G.* By supposition there exists a ∇-cycle z^r on K such that $z'_A=J^K_A z^r$ and therefore (we write E instead of E^A_K)

$$\begin{aligned} z^r &= Ez^r_A + J^K_G z^r, \\ 0 = \nabla z^r &= \nabla E z^r_A + \nabla J^K_G z^r, \\ \nabla E z^r_A &= -\nabla J^K_G z^r. \end{aligned}$$

As $J_G^K z^r$ is a chain on G, $\nabla E z_A^r$ bounds on G and our assertion is proved.

Thus *the operator* ∇E *produces a homomorphism of* $\nabla_{A:K}^r$ *into* $\nabla_{G:K}^{r+1}$. We call this homomorphism again ∇E and prove that it is an isomorphism of $\nabla_{A:K}^r$ onto $\nabla_{G:K}^{r+1}$.

To prove that ∇E maps $\nabla_{A:K}^r$ on the whole $\nabla_{G:K}^{r+1}$ take arbitrarily

$$z_G^{r+1} \in \mathfrak{z}_G^{r+1} \in \nabla_{G:K}^{r+1}.$$

We have to find such a ∇-cycle z_A^r on A that $z_G^{r+1} - \nabla E z_A^r$ bounds on G.

By supposition there exists a chain x^r on K bounded by z_G^{r+1} and consequently

$$\nabla_A J_A^K x^r = J_A^K \nabla x^r = J_A^K z_G^{r+1} = 0;$$

hence $z_A^r = J_A^K x^r$ is a ∇-cycle on A and $x_G^r = x^r - E J_A^K x^r$ is a chain on G with

$$\nabla x_G^r = \nabla x^r - \nabla E J_A^K x^r = z_G^{r+1} - \nabla E z_A^r. \tag{15.1}$$

It remains to be proved that ∇E maps $\nabla_{A:K}^r$ on $\nabla_{G:K}^{r+1}$ *isomorphically*. To this end we show that if for a certain

$$z_A^r \in \mathfrak{z}_A^r \in \nabla_{A:K}^r$$

the ∇-cycle $\nabla E z_A^r$ bounds on G then there exists such a ∇-cycle z^r on K that $z_A^r = J_A^K z^r$.

By supposition

$$\nabla E z_A^r = \nabla x_G^r, \qquad x_G^r \in L_G^r.$$

Therefore

$$\nabla(E z_A^r - x_G^r) = 0,$$

so that $z^r = E z_A^r - x_G^r$ is a ∇-cycle on K with

$$J_A^K z^r = J_A^K E z_A^r - J_A^K x_G^r = z_A^r,$$

q.e.d.

REMARK. In an analogous way to any Δ-cycle z_G^{r+1} on G corresponds the Δ-cycle $\Delta E_K^G z_G^{r+1}$ on A bounding on K, and the operator ΔE_K^G produces an isomorphism of the group $\Delta_{G:K}^{r+1}$ onto the group $\Delta_{A:K}^r$.

§4. SPECIAL CASES

16. Numerical relations. In this section the coefficient-group is either the field I_m, m being a prime, or the field $\mathfrak{R}$ of the rational numbers, the latter being treated according to the remark of section 5. Now *all dualities turn into isomorphisms.*

Let ρ denote the rank of a group of linear forms with respect to the given coefficient-field. We put $\pi_K^r = \rho \Delta_K^r = \rho \nabla_K^r$ (Betti numbers) and

$$\overset{r}{\pi}_{KA} = \rho\overset{r}{\nabla}_{AK} = \rho\overset{r}{\Delta}_{KA}; \qquad \overset{r}{\pi}_{KG} = \rho\overset{r}{\Delta}_{GK} = \rho\overset{r}{\nabla}_{KG},$$
$$\overset{r}{\pi}_{A:K} = \rho\overset{r}{\Delta}_{A:K} = \rho\overset{r}{\nabla}_{A:K}; \qquad \overset{r}{\pi}_{G:K} = \rho\overset{r}{\nabla}_{G:K} = \rho\overset{r}{\Delta}_{G:K}.$$

Then by the first pair of isomorphisms

$$\overset{r}{\pi}_{A} = \overset{r}{\pi}_{A:K} + \overset{r}{\pi}_{KA}; \qquad \overset{r}{\pi}_{G} = \overset{r}{\pi}_{G:K} + \overset{r}{\pi}_{KG}, \tag{16.1}$$

while by the second pair of isomorphisms

$$\overset{r}{\pi}_{K} = \overset{r}{\pi}_{KA} + \overset{r}{\pi}_{KG} \tag{16.2}$$

or

$$\overset{r}{\pi}_{K} = (\overset{r}{\pi}_{A} - \overset{r}{\pi}_{A:K}) + (\overset{r}{\pi}_{G} - \overset{r}{\pi}_{G:K}).$$

Using the third pair of isomorphisms and changing r into $r+1$ we get

$$\overset{r+1}{\pi}_{G} = \overset{r}{\pi}_{A:K} + \overset{r+1}{\pi}_{K} - (\overset{r+1}{\pi}_{A} - \overset{r+1}{\pi}_{A:K}), \tag{16.3}$$

or

$$\overset{r+1}{\pi}_{G} = \overset{r}{\pi}_{A:K} + (\overset{r+1}{\pi}_{K} - \overset{r+1}{\pi}_{KA}). \tag{16.4}$$

In the special case when K is an n-dimensional (orientable) manifold an equivalent formula has been given by Pontrjagin; in this case by the Poincaré duality (see section 42 below) we have $\pi_G^{r+1} = \pi_G^{n-r-1}$ and (16.4) can be written as

$$\overset{n-r-1}{\pi}_{G} = \overset{r}{\pi}_{A:K} + (\overset{r+1}{\pi}_{K} - \overset{r+1}{\pi}_{KA}),$$

which is the Pontrjagin formula [11, p. 449, Theorem II].

17. **Application to regular components.** Now suppose that K is a *homogeneously n-dimensional* cell complex (that is, there are in K no cells of a dimension greater than n while each cell of a dimension less than n is a face of an n-dimensional cell). An $(n-1)$-cell t^{n-1} is called *regular* if there are in K two cells t_1^n and t_2^n such that t^{n-1} is a face of $\pm t_1^n$ and $\pm t_2^n$ and of no more cells of K. A (finite) sequence $t_1^n, \cdots, t_s^n$ of n-cells is called an *r-sequence* if in it any two subsequent cells have in K a common face of a dimension at least $r-1$. Two n-cells of K belong (with all their faces) to the same *r-component* of K if they can be connected (in K) by an r-sequence. The cells common to two different r-components of K are of dimension at most $r-2$. A complex is called r-connected if it consists only of one r-component; an n-connected complex K is called an n-dimensional *pseudomanifold* if all of its $(n-1)$-cells are regular. There are orientable and non-orientable n-dimensional pseudomanifolds: the group $\nabla_K^n(I)$ of an n-dimensional pseudomanifold K is always cyclic, infinite if K is orientable, and of order 2 if K is non-orientable. It follows that if the n-dimensional cell complex K is in any way decomposed into a certain number of n-dimensional pseudomanifolds no two of which

have a cell of dimension at least $n-1$ in common, then the group $\nabla_K^n(I)$ is the direct sum of as many infinite cyclic groups as there are orientable pseudomanifolds and of as many groups of order 2 as there are non-orientable pseudomanifolds in our decomposition.

Let K be again an arbitrary homogeneously n-dimensional complex; denote by A_{n-1} the subcomplex of K consisting of all $(n-1)$-dimensional irregular cells and of all faces of these cells. The n-components of the open subcomplex $G_{n-1}=K\setminus A_{n-1}$ are called *regular components of* K (we slightly change here the terminology of [4, p. 190]). The decomposition of G_{n-1} into its n-components (the regular components of K) is a decomposition of the kind just mentioned and therefore *the number* $q(K)$ *of the regular components of* K *is equal to the* n*-dimensional Betti number modulo* 2 *of* G_{n-1}, *while the number* $q_o(K)$ *of the orientable regular components of* K *is the ordinary* n*-dimensional Betti number of* G_{n-1}. Denoting as usual by π^r the ordinary Betti numbers and by $\pi^r(2)$ the Betti numbers modulo 2 we get, using (16.4):

17.1. For any homogeneously n-dimensional complex K:

$$(17.1)\qquad q(K)=\pi^{n-1}_{A_{n-1}:K}(2)+\pi^n_K(2);\qquad q_o(K)=\pi^{n-1}_{A_{n-1}:K}+\pi^r_K.$$

In particular, if all $(n-1)$-cells of K are regular, then

$$(17.11)\qquad q(K)=\pi^n_K(2);\qquad q_o(K)=\pi^n_K.$$

This is the case with the complex $G=K\setminus A$ if the closed subcomplex $A\subset K$ contains all irregular $(n-1)$-cells of K. Therefore, applying again (16.4), we get:

17.2. If the closed subcomplex A of the homogeneously n-dimensional complex K contains all irregular $(n-1)$-cells of K, then

$$(17.2)\qquad q(G)=\pi^{n-1}_{A:K}(2)+\pi^n_K(2)-\pi^n_{KA}(2);\qquad q_o(G)=\pi^{n-1}_{A:K}+\pi^n_K-\pi^n_{KA}.$$

If A is moreover $(n-1)$-dimensional, then

$$(17.3)\qquad q(G)=\pi^{n-1}_{A:K}(2)+\pi^n_K(2);\qquad q_o(G)=\pi^{n-1}_{A:K}+\pi^n_K.$$

Now we call *any cell* of K *regular* if the n-dimensional ∇^n-group of its open star (with respect to the group I) is infinite cyclic. This definition agrees for $(n-1)$-cells with the previous one. If A contains all irregular cells of dimensions at least r then the r-components of G are identical with the regular components of G. Thus for $r=0$, that is *if* A *contains all irregular cells of* K, *the ordinary components of* G *coincide with the regular ones and their number is still given by* (17.2) *and, in the case of an* $(n-1)$*-dimensional* A, *by* (17.3).

REMARK. If no special hypotheses on the closed subcomplex A are made, then the number of components of the open subcomplex $G=K\setminus A$ depends upon the homological properties of the mutual situation in K of two subcomplexes: the given subcomplex A and the subcomplex A_0 of all irregular ele-

ments of K. Even the most elementary examples show that the features of this mutual situation may be very different.

18. **The Phragmén-Brouwer theorem.** Let A be as always a closed subcomplex of the cell complex K, $G=K\backslash A$ and

$$A = A_1 \cup A_2, \quad A_1 \text{ and } A_2 \text{ closed in } A,$$
$$Q_1 = A\backslash A_1, \quad Q_2 = A\backslash A_2; \qquad G_1 = K\backslash A_1, \quad G_2 = K\backslash A_2,$$
$$A_0 = A_1 \cap A_2.$$

Denote finally by $E^i_{A:K}$, $i=1, 2$, the homomorphism of $\nabla^r_{Q_i}$ into $\nabla^r_{A:K}$ which assigns to each element z^r of $\nabla^r_{Q_i}$ the class $\mathfrak{z}^r \in \nabla^r_{A:K}$ containing the element $E^{Q_i}_A z^r$. We shall prove the isomorphism

$$(E^1_{A:K}\nabla^r_{Q_1}) \cap (E^2_{A:K}\nabla^r_{Q_2}) \approx \nabla^{r+1}_{G:G_1} \cap \nabla^{r+1}_{G:G_2}. \tag{18.1}$$

Note that *the elements of the group* $(E^1_{A:K}\nabla^r_{Q_1})\cap(E^2_{A:K}\nabla^r_{Q_2})$ *can be defined as being those classes* $\mathfrak{z}^r \in \nabla^r_{A:K}$ *of* ∇*-cycles which contain both cycles on* Q_1 *and* Q_2; the elements of $\nabla^r_{G:G_1}\cap\nabla^r_{G:G_2}$ are ∇-classes of G whose elements bound both on G_1 and G_2.

The isomorphism (18.1) is realized by the operator ∇E^A_K: let $\mathfrak{z}^r_0 \in (E^1_{A:K}\nabla^r_{Q_1}) \cap (E^2_{A:K}\nabla^r_{Q_2})$. Take $z^r_0 \in \mathfrak{z}^r_0$ on Q_1. Then $E^A_K z^r_0$ is equal to zero on A_1, thus $\nabla E^A_K z^r_0$ bounds on G_1. Consequently *all* elements of the ∇-class $\nabla E^A_K \mathfrak{z}^r_0 \in \nabla^{r+1}_G$ bound on G_1. Taking $z^r_0 \in \mathfrak{z}^r_0$ on Q_2 we see in the same way that all elements of $\nabla E^A_K \mathfrak{z}^r_0$ bound on G_2. Hence the isomorphism ∇E^A_K maps the left-hand member of (18.1) into the right-hand one. To show that this mapping is *on* the right-hand member of (18.1) let

$$z^{r+1}_G \in E^A_K \mathfrak{z}^r_0 = \mathfrak{z}^{r+1}_G \in \nabla^{r+1}_{G:G_1} \cap \nabla^{r+1}_{G:G_2}.$$

As z^{r+1}_G bounds on G_i, $i=1, 2$, it is $z^{r+1}_G = \nabla x^r_i$ with x^r_i on G_i. Therefore $z^r_A = J^K_A x^r_i$ is a ∇-cycle on Q_i and by (15.1) (with $x^r_G = x^r_i - EJ^K_A x^r_i$) we have:

$$\nabla E^A_K z^r_A \in \mathfrak{z}^{r+1}_G,$$

thus $z^r_A \in \mathfrak{z}^r_0$, and $\mathfrak{z}^r_0$ contains cycles on Q_1 as well as cycles on Q_2, q.e.d.

Let K *be simply connected with respect to* r; then the isomorphism (18.1) is replaced by

$$\nabla^r_{AQ_1} \cap \nabla^r_{AQ_2} \approx \nabla^{r+1}_{G:G_1} \cap \nabla^{r+1}_{G:G_2}. \tag{18.2}$$

Now

$$\Delta^r_A \supseteq \Delta^r_{AA_i} \perp \nabla^r_{AQ_i} \subseteq \nabla^r_A, \qquad i = 1, 2,$$

therefore

$$\Delta^r_A \supseteq (\Delta^r_{AA_1} + \Delta^r_{AA_2}) \perp (\nabla^r_{AQ_1} \cap \nabla^r_{AQ_2}) \subseteq \nabla^r_A$$

(the sum at left may fail to be direct) and

$$(18.3) \qquad (\Delta^r_A - (\Delta^r_{AA_1} + \Delta^r_{AA_2})) \mid (\nabla^r_{AQ_1} \cap \nabla^r_{AQ_2}).$$

To any Δ-cycle z^r on A corresponds the Δ-cycle

$$\Delta J^A_{Q_1} z^r = -\Delta J^A_{A_1} z^r \text{ on } A_0 = A_1 \cap A_2$$

(where Δ operates on A). This correspondence generates a homomorphism (the "Naht-Homomorphismus" of [4, p. 292]) of Δ^r_A onto $\Delta^{r-1}_{A_0:A_1} \cap \Delta^{r-1}_{A_0:A_2}$ with the kernel $\Delta^r_{AA_1} + \Delta^r_{AA_2}$ (the proof is easy and is given in [4, pp. 289–293]). Thus

$$(18.4) \qquad (\Delta^r_A - (\Delta^r_{AA_1} + \Delta^r_{AA_2})) \approx \Delta^{r-1}_{A_0:A_1} \cap \Delta^{r-1}_{A_0:A_2}.$$

From (18.2)–(18.4) follows a duality which we call

THE PHRAGMÉN-BROUWER THEOREM.

$$(18.5) \qquad (\Delta^{r-1}_{A_0:A_1} \cap \Delta^{r-1}_{A_0:A_2}) \mid (\nabla^{r+1}_{G:G_1} \cap \nabla^{r+1}_{G:G_2}).$$

19. **Examples.**

1. The elements of the complex K are: a triangle, its edges and its vertices, thus the cells $\pm t^2$, $\pm t^1_i$, $\pm t^0_i$, $i=1, 2, 3$, with obvious incidence coefficients. The open subcomplex G consists only of $\pm t^2$, the closed subcomplex A of the remaining elements of K. The operator ∇E^A_K assigns to the ∇-class $z^1_A \in \nabla^1_A$ consisting of the ∇-cycles t^1_1, t^2_2, t^1_3 on A the ∇-class $z^2_G \in \nabla^2_G$ the only element of which is t^2. The groups $\nabla^1_A(I)$ and $\nabla^2_G(I)$ are both infinite cyclic (duality theorem of Alexander-Kolmogoroff).

2. Consider the circular ring in the plane decomposed into the cell complex K whose elements are: $\pm t^2$, $\pm t^1_i$, $\pm t^0_i$, $i=1, 2, 3$, with the incidence coefficients

$$(t^2 : t^1_1) = (t^2 : t^1_2) = 1, \qquad (t^2 : t^1_3) = 0, \qquad (t^1_3 : t^0_1) = 1, \qquad (t^1 : t^0_2) = -1,$$

$$(t^1_i : t^0_j) = 0 \quad \text{if} \quad i, j = 1, 2.$$

The open subcomplex G contains only the cells $\pm t^2$ and $A = K \backslash G$. All groups of the figure K, A, G are determined by their ranks:

$$\pi^0_K = \pi^0_{KA} = 1; \quad \pi^0_A = 2; \quad \pi^0_{A:K} = 1; \quad \pi^0_G = \pi^0_{KG} = \pi^0_{G:K} = 0,$$

$$\pi^1_K = \pi^1_{KA} = 1; \quad \pi^1_A = 2; \quad \pi^1_{A:K} = 1; \quad \pi^1_G = \pi^1_{G:K} = 1; \quad \pi^1_{KG} = 0,$$

$$\pi^2_K = \pi^2_{KA} = \pi^2_A = \pi^2_{A:K} = \pi^2_{KG} = 0; \quad \pi^2_G = \pi^2_{G:K} = 1.$$

3. The elements of K are $\pm t^2$, $\pm t^1$, $\pm t^0$ with the incidence coefficients $(t^2 : t^1) = 2$, $(t^1 : t^0) = 0$; the elements of A are $\pm t^1$ and $\pm t^0$ (the cell complex K is a cell decomposition of the projective plane; then A is a line on this plane). The group $\Delta^1_{A:K}$ consists of all elements of the form $\pm 2nt^1$ with an arbitrary

integer n, while the elements of $\Delta^2_{G:K}=\Delta^2_G$ are of the form $n\iota^2$; thus $\Delta^1_{A:K}$ and $\Delta^2_{G:K}$ are both infinite cyclic.

4. The cells of K are: $\pm\iota^2_1$, $\pm\iota^2_2$; $\pm\iota^1$; $\pm\iota^0$ with the incidence coefficients

$$(\iota^2_1:\iota^1)=(\iota^2_2:\iota^1)=2; \qquad (\iota^1:\iota^0)=0$$

(two projective planes intersecting in a line); let the cells of A be $\pm\iota^2_1$, $\pm\iota^1$, $\pm\iota^0$. The coefficient-groups are I for ∇ and $\bar{I}$ for Δ. Then ∇^2_K is the direct sum of an infinite cyclic group (generated by ι^2_1) and of a group of order 2 (generated by $\iota^2_1-\iota^2_2$). Correspondingly $\Delta^2_K\approx\bar{I}+I_2$ and

$$\overset{2}{\nabla}_{KG}=\overset{2}{\nabla}_G\approx I, \qquad \overset{2}{\Delta}_{KA}\approx I_2, \qquad \overset{2}{\nabla}_{AK}=\overset{2}{\nabla}_A\approx I_2.$$

The geometric meaning of the annihilation $\Delta^2_K\supseteq\Delta^2_{KA}\perp\nabla^2_{KG}\subseteq\nabla^2_K$ is in this case particularly simple.

Now let K be the same, while the cells of A are $\pm\iota^1$, $\pm\iota^0$ (that is A is the intersection line of the two planes). Then $\nabla^1_{A:K}=\nabla^1_A\approx I$, $\nabla^2_{KG}=\nabla^2_K\approx I+I_2$, $\nabla^2_G\approx I+I$, while $\nabla^2_{G:K}$ (isomorphic with $\nabla^1_{A:K}$) is the infinite cyclic group generated by $2(\iota^2_1-\iota^2_2)$.

5. Let $\mathbf{K}$ be the three-dimensional (continuous) polyhedron composed by the interior and the surface of an ordinary anchor ring. Let $\mathbf{A}$ be a ring-shaped body (homeomorphic with $\mathbf{K}$) lying in the interior of $\mathbf{K}$ and going twice around $\mathbf{K}$. We suppose moreover that $\mathbf{A}$ is decomposed into the simplices of a certain closed subcomplex A of a simplicial decomposition K of $\mathbf{K}$; then $\Gamma=\mathbf{K}\setminus\mathbf{A}$ is decomposed into the simplices of the open subcomplex $G=K\setminus A$ of K. The coefficient-groups are I for Δ and $\bar{I}$ for ∇. Then $\Delta^1_{KA}=2\Delta^1_K$, $\Delta^1_{GK}\approx\nabla^1_{KG}\approx I_2$ (the equator of K generates the group Δ^1_{GK}; taken twice it bounds on G). The group Δ^1_{KA} is not a direct summand of the group Δ^1_K.

6. Let K be a complex which arises when we identify the two poles of an ordinary two-dimensional sphere. A is composed of two parallel circles of this sphere. It follows from (16.4):

$$\overset{2}{\pi}_G=\overset{1}{\pi}_{A:K}+\overset{2}{\pi}_K-\overset{2}{\pi}_{KA}=2+1-0=3.$$

7. In addition to the well known elementary examples illustrating the theorem of Phragmén-Brouwer we illustrate the formula (18.1) by a torus surface K on which two meridians A_1 and A_2 form the complex A; the whole figure is considered in a cell decomposition in which the two meridians A_1 and A_2, oriented in opposite senses, are cells ι^1_1 and ι^1_2. Then $\iota^1_1-\iota^1_2$ is a ∇-cycle on A extensible over K; therefore ι^1_1 and ι^1_2 belong to the same class $\mathfrak{z}^1\in\nabla^1_{A:K}$; the group $(E^1_{A:K}\nabla^1_{Q_1})\cap(E^2_{A:K}\nabla^1_{Q_2})$ is generated by this $\mathfrak{z}^1$ and is infinite cyclic. The group $(\nabla^2_{G:G_1}\cap\nabla^2_{G:G_2})\approx I$ is generated by $\iota^2_1-\iota^2_2$, where the cells ι^2_1 and ι^2_2 correspond to the two domains into which the ring surface is decomposed by the meridians A_1 and A_2. An open plane circular ring K and two radial segments A_1 and A_2 on it could be treated in the same way.

CHAPTER II. SPECTRA

§5. CELL SPECTRA

20. **Projections.** Let K_α and K_β be two cell complexes. Let there be given for each dimension r a homomorphism ϖ_α^β of the group $L_\beta^r(I)$ of all r-dimensional integer chains of K_β into the group $L_\alpha^r(I)$ of all r-dimensional integer chains of K_α. We suppose moreover that these homomorphisms commute with the operator Δ, that is, that for any (integer) chain x_β^r on K_β

$$\Delta \overset{\beta}{\varpi}_\alpha \overset{r}{x}_\beta = \overset{\beta}{\varpi}_\alpha \Delta \overset{r}{x}_\beta. \tag{20.1}$$

Under these conditions the system of the given homomorphisms ϖ_α^β is called a *projection* of the cell complex K_β into the cell complex K_α.

Let $A_\alpha \subseteq K_\alpha$, $A_\beta \subseteq K_\beta$ be closed subcomplexes of K_α, K_β respectively. We put $G_\alpha = K_\alpha \backslash A_\alpha$, $G_\beta = K_\beta \backslash A_\beta$ and write systematically L_α^r, $L_{\alpha 0}^r$, $L_{\alpha 1}^r$ for $L_{K_\alpha}^r$, $L_{A_\alpha}^r$, $L_{G_\alpha}^r$, $\cdots$, and $J_{\alpha 0}^\alpha$, $J_{\alpha 1}^\alpha$, $\cdots$ for $J_{A_\alpha}^{K_\alpha}$, $J_{G_\alpha}^{K_\alpha}$, $\cdots$.

Any chain on A_α, A_β being a chain on K_α, K_β, we say that the projection ϖ_α^β is *cogredient* (with A_β, A_α) if every (integer) chain of A_β is mapped by ϖ_α^β onto a chain on A_α. As for x_β^r on A_β, x_α^r on A_α, the boundaries Δx_β^r, Δx_α^r are the same on A_β, A_α and on K_β, K_α, the projection ϖ_α^β of K_β into K_α (cogredient with A_β, A_α) generates a projection $\varpi_{\alpha 0}^{\beta 0}$ of A_β into A_α.

Now define for any chain $x_{\beta 1}^r$ on G_β:

$$\overset{\beta 1}{\varpi}_{\alpha 1} \overset{r}{x}_{\beta 1} = J_{\alpha 1}^\alpha \overset{\beta}{\varpi}_\alpha \overset{r}{x}_{\beta 1}. \tag{20.11}$$

Then whatever be the chain x_β on K_β, $J_{\alpha 1}^\alpha \varpi_\alpha^\beta x_\beta = J_{\alpha 1}^\alpha \varpi_\alpha^\beta J_{\beta 1}^\beta x_\beta$. We use this remark in

$$\begin{aligned}\Delta_{\alpha 1} \overset{\beta 1}{\varpi}_{\alpha 1} \overset{r}{x}_{\beta 1} &= \Delta_{\alpha 1} J_{\alpha 1}^\alpha \overset{\beta}{\varpi}_\alpha \overset{r}{x}_{\beta 1} = J_{\alpha 1}^\alpha \Delta_\alpha \overset{\beta}{\varpi}_\alpha \overset{r}{x}_{\beta 1} = J_{\alpha 1}^\alpha \overset{\beta}{\varpi}_\alpha \Delta_\beta \overset{r}{x}_{\beta 1} = J_{\alpha 1}^\alpha \overset{\beta}{\varpi}_\alpha J_{\beta 1}^\beta \Delta_\beta \overset{r}{x}_{\beta 1} \\ &= \overset{\beta 1}{\varpi}_{\alpha 1} J_{\beta 1}^\beta \Delta_\beta \overset{r}{x}_{\beta 1} = \overset{\beta 1}{\varpi}_{\alpha 1} \Delta_{\beta 1} \overset{r}{x}_{\beta 1},\end{aligned}$$

where Δ_α, Δ_β, $\Delta_{\alpha 1}$, $\Delta_{\beta 1}$ operate on K_α, K_β, G_α, G_β.

20.2. Thus a projection ϖ_α^β of K_β into K_α cogredient with A_β, A_α generates a projection $\varpi_{\alpha 0}^{\beta 0}$ of A_β into A_α (identical on A_β with ϖ_α^β) and (by (20.11)) a projection $\varpi_{\alpha 1}^{\beta 1}$ of G_β into G_α.

21. **Conjugate homomorphisms.** The projection ϖ_α^β of K_β into K_α assigns to each chain $x_\beta^r = \sum a_j i_{\beta j}^r$ on K_β (with respect to a given coefficient-group $\overline{\mathfrak{A}}$) the chain $\varpi_\alpha^\beta x_\beta^r = \sum_j a_j \varpi_\alpha^\beta i_{\beta j}^r$ on K (with respect to the same coefficient-group $\overline{\mathfrak{A}}$) and thus produces for any r a homomorphism ϖ_α^β of the group $\overline{L}_\beta^r = L_\beta^r(\overline{\mathfrak{A}})$ into $\overline{L}_\alpha^r = L^r(\overline{\mathfrak{A}})$ which will still be called a projection. The conjugate homomorphism π_β^α of $L_\alpha^r = L(\mathfrak{A})$ into $L_\beta^r = L_\beta^r(\mathfrak{A})$, where $\mathfrak{A} \mid \overline{\mathfrak{A}}$, assigns to any chain $x_\alpha^r \in L_\alpha^r$ the chain $\pi_\beta^\alpha x_\alpha^r \in L_\beta^r$ defined by

$$(\overset{\alpha}{\pi}_\beta \overset{r}{x}_\alpha \cdot i_\beta^r) = (x_\alpha^r \cdot \overset{\beta}{\varpi}_\alpha i_\beta^r), \tag{21.1}$$

whatever $i_\beta^r \in K_\beta$. The homomorphisms π_β^α commute with ∇ as is seen by

$$(\nabla\pi_\beta^\alpha x_\alpha^r \cdot t_\beta^{r+1}) = (\pi_\beta^\alpha x_\alpha^r \cdot \Delta t_\beta^{r+1}) = (x_\alpha^r \cdot \varpi_\alpha^\beta \Delta t_\beta^{r+1}) = (x_\alpha^r \cdot \Delta\varpi_\alpha^\beta t_\beta^{r+1})$$
$$= (\nabla x_\alpha^r \cdot \varpi_\alpha^\beta t_\beta^{r+1}) = (\pi_\beta^\alpha \nabla x_\alpha^r \cdot t_\beta^{r+1}).$$

Moreover, if ϖ_α^β is cogredient with A_β, A_α then for a chain $x_{\alpha 1}^r$ on G_α and any $t_\beta^r \in A_\beta$,

$$(\pi_\beta^\alpha x_{\alpha 1}^r \cdot t_\beta^r) = (x_{\alpha 1}^r \cdot \varpi_\alpha^\beta t_\beta^r) = 0,$$

so that π_β^α maps $L_{\alpha 1}^r$ into $L_{\beta 1}^r$ and thus produces the homomorphism $\pi_{\beta 1}^{\alpha 1}$ of $L_{\alpha 1}^r$ into $L_{\beta 1}^r$ identical with π_β^α on $L_{\alpha 1}^r \subseteq L_\alpha^r$ and conjugate with $\varpi_{\alpha 1}^{\beta 1}$.

On the other hand we have the homomorphism $\pi_{\beta 0}^{\alpha 0}$ conjugate with $\varpi_{\alpha 0}^{\beta 0}$ in the form

$$\pi_{\beta 0}^{\alpha 0} x_{\alpha 0}^r = J_{\beta 0}^\beta \pi_\beta^\alpha x_{\alpha 0}^r \qquad \text{for any } x_{\alpha 0}^r \in L_{\alpha 0}^r.$$

We collect all we have just said about the cogredient projections ϖ_α^β in the following formulas:

$$(21.2)\qquad \pi_\beta^\alpha \left|\begin{matrix} L_\alpha^r & \overline{L}_\alpha^r \\ L_\beta^r & \overline{L}_\beta^r \end{matrix}\right| \varpi_\alpha^\beta; \qquad \pi_{\beta 0}^{\alpha 0} \left|\begin{matrix} L_{\alpha 0}^r & \overline{L}_{\alpha 0}^r \\ L_{\beta 0}^r & \overline{L}_{\beta 0}^r \end{matrix}\right| \varpi_{\alpha 0}^{\beta 0}; \qquad \pi_{\beta 1}^{\alpha 1} \left|\begin{matrix} L_{\alpha 1}^r & \overline{L}_{\alpha 1}^r \\ L_{\beta 1}^r & \overline{L}_{\beta 1}^r \end{matrix}\right| \varpi_{\alpha 1}^{\beta 1}$$

where

$$(21.3)\qquad \begin{cases} \varpi_{\alpha 0}^{\beta 0} = \varpi_\alpha^\beta; & \pi_{\beta 0}^{\alpha 0} = J_{\beta 0}^\beta \pi_\beta^\alpha, \\ \varpi_{\alpha 1}^{\beta 1} = J_{\alpha 1}^\alpha \varpi_\alpha^\beta; & \pi_{\beta 1}^{\alpha 1} = \pi_\beta^\alpha, \end{cases}$$

and ϖ_α^β, $\varpi_{\alpha 0}^{\beta 0}$, $\varpi_{\alpha 1}^{\beta 1}$ commute with Δ, while π_β^α, $\pi_{\beta 0}^{\alpha 0}$, $\pi_{\beta 1}^{\alpha 1}$ commute with ∇. Therefore the homomorphisms ϖ_α^β, $\varpi_{\alpha 0}^{\beta 0}$, $\varpi_{\alpha 1}^{\beta 1}$ map Δ_β^r, $\Delta_{\beta 0}^r$, $\Delta_{\beta 1}^r$ respectively into Δ_α^r, $\Delta_{\alpha 0}^r$, $\Delta_{\alpha 1}^r$, while π_β^α, $\pi_{\beta 0}^{\alpha 0}$, $\pi_{\beta 1}^{\alpha 1}$ map ∇_α^r, $\nabla_{\alpha 0}^r$, $\nabla_{\alpha 1}^r$ into ∇_β^r, $\nabla_{\beta 0}^r$, $\nabla_{\beta 1}^r$:

$$(21.4)\qquad \pi_\beta^\alpha \left|\begin{matrix} \nabla_\alpha^r & \Delta_\alpha^r \\ \nabla_\beta^r & \Delta_\beta^r \end{matrix}\right| \varpi_\alpha^\beta; \qquad \pi_{\beta 0}^{\alpha 0} \left|\begin{matrix} \nabla_{\alpha 0}^r & \Delta_{\alpha 0}^r \\ \nabla_{\beta 0}^r & \Delta_{\beta 0}^r \end{matrix}\right| \varpi_{\alpha 0}^{\beta 0}; \qquad \pi_{\beta 1}^{\alpha 1} \left|\begin{matrix} \nabla_{\alpha 1}^r & \Delta_{\alpha 1}^r \\ \nabla_{\beta 1}^r & \Delta_{\beta 1}^r \end{matrix}\right| \varpi_{\alpha 1}^{\beta 1}.$$

From the definition of $\varpi_{\alpha 0}^{\beta 0}$, $\varpi_{\alpha 1}^{\beta 1}$, $\cdots$ follows furthermore:

$$(21.5)\qquad \varpi_\alpha^\beta E_\beta^{\beta 0} \tilde{x}_{\beta 0}^r = E_\alpha^{\alpha 0} \varpi_{\alpha 0}^{\beta 0} \tilde{x}_{\beta 0}^r; \qquad \pi_\beta^\alpha E_\alpha^{\alpha 1} x_{\alpha 1}^r = E_\beta^{\beta 1} \pi_{\beta 1}^{\alpha 1} x_{\alpha 1}^r,$$

$$(21.6)\qquad \pi_{\beta 0}^{\alpha 0} J_{\alpha 0}^\alpha x_\alpha^r = J_{\beta 0}^\beta \pi_\beta^\alpha x_\alpha^r; \qquad \varpi_{\alpha 1}^{\beta 1} J_{\beta 1}^\beta \tilde{x}_\beta^r = J_{\alpha 1}^\alpha \varpi_\alpha^\beta \tilde{x}_\beta^r.$$

22. **Cell spectra.** Two projections ϖ_α^β and $\varpi_\alpha'^\beta$ of K_β into K_α are called *homologous* to each other if they produce the same homomorphism of Δ_β^r into Δ_α^r.

Now let there be given an unbounded partially ordered set of cell complexes K_α; we write $\beta > \alpha$ if in this set K_β follows K_α; "unbounded" means that to each two elements of the set a third element of the same set can be found which follows on each of the two given elements.

Suppose that for any $\beta > \alpha$ a finite number of "allowed" projections ϖ_α^β of K_β into K_α is given in such a manner that:

1. Any two allowed projections of K_β into K_α are homologous.

2. If $\gamma>\beta>\alpha$ and ϖ_β^γ, ϖ_α^β are allowed projections then the projection $\varpi_\alpha^\gamma=\varpi_\alpha^\beta(\varpi_\beta^\gamma)$ of K_γ into K_α is allowed.

Under these hypotheses the partially ordered set of the cell complexes K_α with their allowed projections ϖ_α^β is called a *cell spectrum* and denoted by

$$K = \{K_\alpha, \varpi_\alpha^\beta\}. \tag{22}$$

A cell spectrum (22) determines two *group spectra* [1, pp. 58–62] or *homomorphism systems* [13, pp. 668–669]: the *inverse* group spectrum $(\Delta_\alpha^r, \varpi_\alpha^\beta)$ and the *direct* group spectrum $(\nabla_\alpha^r, \pi_\beta^\alpha)$, the homomorphisms of the two spectra being conjugate. Therefore the limit groups (see section 23) called the *Betti groups of the cell spectrum* K,

$$\Delta_K^r = \liminf (\Delta_\alpha^r, \varpi_\alpha^\beta); \qquad \nabla_K^r = \limsup (\nabla_\alpha^r, \pi_\beta^\alpha),$$

are dual.

Now let the closed subcomplexes $A_\alpha\subseteq K_\alpha$ be given in such a way that

1. each allowed projection ϖ_α^β is cogredient with A_β, A_α;

2. each two projections $\varpi_{\alpha 0}^{\beta 0}$, $\varpi_{\alpha 0}^{\prime\beta 0}$ of A_β into A_α; $\varpi_{\alpha 1}^{\beta 1}$, $\varpi_{\alpha 1}^{\prime\beta 1}$ of G_β into G_α, generated by allowed projections ϖ_α^β, $\varpi_\alpha^{\prime\beta}$ of K_β into K_α, are homologous to each other. Then the given cell spectrum (22) is called *cogredient with the subcomplexes* A_α *and* G_α; such a spectrum defines the cell spectra

$$A = \{A_\alpha, \varpi_{\alpha 0}^{\beta 0}\}; \qquad G = \{G_\alpha, \varpi_{\alpha 1}^{\beta 1}\}$$

with their Betti groups

$$\Delta_A^r = \liminf (\Delta_{\alpha 0}^r, \varpi_{\alpha 0}^{\beta 0}), \qquad \nabla_A^r = \limsup (\nabla_{\alpha 0}^r, \pi_{\beta 1}^{\alpha 0});$$

$$\Delta_G^r = \liminf (\Delta_{\alpha 1}^r, \varpi_{\alpha 1}^{\beta 1}), \qquad \nabla_G^r = \limsup (\nabla_{\alpha 1}^r, \pi_{\beta 1}^{\alpha 1})$$

and the dualities

$$\Delta_A^r \mid \nabla_A^r; \qquad \Delta_G^r \mid \nabla_G^r.$$

§6. The extension and the intersection homomorphisms and the groups of the figure K, A, G for cell spectra

23. Lemmas on group spectra. In this section "spectrum" means "group spectrum." Let $(U_\alpha, \pi_\beta^\alpha)$ be a direct spectrum with the limit group U. The set theoretical sum of all groups U_α is decomposed into *bundles* or classes of equivalent elements, where two elements $u_\alpha\in U_\alpha$ and $u_\beta\in U_\beta$ are called equivalent if there exists a $\gamma>\alpha$, β such that $\pi_\gamma^\alpha u_\alpha=\pi_\gamma^\beta u_\beta$. These bundles are the elements of the limit group U. To get the sum of $u=\{u_\alpha\}\in U$ and $u'=\{u_\alpha'\}\in U$ take $u_\alpha\in u$, $u_\beta'\in u'$ and any $\gamma>\alpha$, β. Then $u+u'$ is defined as the bundle containing $\pi_\gamma^\alpha u_\alpha+\pi_\gamma^\beta u_\beta$ (see [1, p. 59]). Let $(\overline{U}_\alpha, \varpi_\alpha^\beta)$ be an in-

verse spectrum. The elements of the limit group $\overline{U}=\lim\inf(\overline{U}_\alpha, \varpi_\alpha^\beta)$ are "threads," that is systems $\hat{u}=\{\hat{u}_\alpha\}$ of elements $\hat{u}_\alpha\in\overline{U}_\alpha$ such that:

1. the thread $\hat{u}$ contains one and only one element $\hat{u}_\alpha$ of each group $\overline{U}_\alpha$;
2. if $\hat{u}_\alpha\in\hat{u}$, $\hat{u}_\beta\in\hat{u}$ and $\beta>\alpha$ then $\hat{u}_\alpha=\varpi_\alpha^\beta\hat{u}_\beta$.

The sum of two threads $\hat{u}=\{\hat{u}_\alpha\}$ and $\hat{u}'=\{\hat{u}_\alpha'\}$ is the thread $\{\hat{u}_\alpha+\hat{u}_\alpha'\}$ (see [1, p. 58]).

LEMMA 23.1 (obvious). *If in an inverse spectrum $(\overline{U}_\alpha, \varpi_\alpha^\beta)$ the subgroups $\overline{U}_{\alpha 0}\subseteq\overline{U}_\alpha$ are given in such a way that $\varpi_\alpha^\beta\overline{U}_{\beta 0}\subseteq\overline{U}_{\alpha 0}$ for $\beta>\alpha$ then the groups $\overline{U}_{\alpha 0}$ with their homomorphisms $\varpi_{\alpha 0}^{\beta 0}=\varpi_\alpha^\beta$ form an inverse spectrum and $\overline{U}_0=\lim\inf$ $(\overline{U}_{\alpha 0}, \varpi_{\alpha 0}^{\beta 0})$ is a subgroup of $\overline{U}=\lim\inf(\overline{U}_\alpha, \varpi_\alpha^\beta)$.*

If in a direct spectrum $(U_\alpha, \pi_\beta^\alpha)$ the subgroups $U_{\alpha 0}\subseteq U_\alpha$ are given in such a way that $\pi_\beta^\alpha U_{\alpha 0}\subseteq U_{\beta 0}$, then the $U_{\alpha 0}$ with their homomorphisms $\pi_{\beta 0}^{\alpha 0}=\pi_\beta^\alpha$ form a direct spectrum $(U_{\alpha 0}, \pi_{\beta 0}^{\alpha 0})$. Any bundle of this spectrum is contained in a bundle of $(U_\alpha, \pi_\beta^\alpha)$ and any bundle of $(U_\alpha, \pi_\beta^\alpha)$ contains at most one bundle of $(U_{\alpha 0}, \pi_{\beta 0}^{\alpha 0})$. Therefore the group $U_0=\lim\sup(U_{\alpha 0}, \pi_{\beta 0}^{\alpha 0})$ is mapped isomorphically on a subgroup of $U=\lim\sup(U_\alpha, \pi_\beta^\alpha)$ and can be identified with this subgroup(3).

LEMMA 23.2. *Under the previous hypotheses the group*

$$\overline{U}_{\beta:\beta 0}=\overline{U}_\beta-\overline{U}_{\beta 0};\qquad U_{\alpha:\alpha 0}=U_\alpha-U_{\alpha 0}$$

is mapped by

$$\varpi_\alpha^\beta \text{ into } \overline{U}_{\alpha:\alpha 0}=\overline{U}_\alpha-\overline{U}_{\alpha 0};\qquad \pi_\beta^\alpha \text{ into } U_{\beta:\beta 0}=U_\beta-U_{\beta 0}$$

and(3)

$$\overline{U}-\overline{U}_0=\lim\inf(\overline{U}_{\alpha:\alpha 0}, \varpi_\alpha^\beta);\qquad U-U_0=\lim\sup(U_{\alpha:\alpha 0}, \pi_\beta^\alpha).$$

Let us sketch the proof of Lemma 23.2. To prove the left-hand assertion we take any element of $\overline{U}-\overline{U}_0$, that is, any class of threads $\hat{u}=\{\hat{u}_\alpha\}\in\overline{U}$ congruent with respect to the subgroup $\overline{U}_0\subseteq\overline{U}$, and assign to each thread $\hat{u}=\{u_\alpha\}$ belonging to this class the thread $\{\mathfrak{u}_\alpha\}$ $\in\lim\inf(\overline{U}_\alpha-\overline{U}_{\alpha 0}, \varpi_\alpha^\beta)$ where $\bar{u}_\alpha\in\bar{\mathfrak{u}}_\alpha\in\overline{U}_\alpha-\overline{U}_{\alpha 0}$. This correspondence is easily seen to be an isomorphism between $\overline{U}-\overline{U}_0$ and $\lim\inf(\overline{U}_\alpha-\overline{U}_{\alpha 0}, \varpi_\alpha^\beta)$.

To prove the right-hand assertion define the group U' as follows. Call the elements u_α and u_β of the set theoretical sum of the groups U_α *equivalent with respect to* U_0 if a $\gamma>\alpha, \beta$ can be found such that $\pi_\gamma^\alpha u_\alpha-\pi_\gamma^\beta u_\beta\in U_{\gamma 0}$. The set theoretical sum of the groups U_α falls now into classes of elements equivalent with respect to U_0 and these classes form the group U' (we define the sum of two classes u and u' as follows: take $u_\alpha\in u$, $u_\beta'\in u'$ and $\gamma>\alpha, \beta$; define $u+u'$

(3) In cases when among all isomorphisms between two isomorphic groups a well defined isomorphism is chosen in some "natural" way we often replace the sign of isomorphism by the equality sign and consider the given groups as being identical.

as the equivalence class containing $\pi_\gamma^\alpha u_\alpha + \pi_\gamma^\beta u_\beta'$). The group U' is easily seen to be isomorphic with both $U - U_0$ and lim sup $(U_\alpha - U_{\alpha 0}, \pi_\beta^\alpha)$.

LEMMA 23.3. *Let two direct spectra* $(U_\alpha, \rho_\beta^\alpha)$ *and* $(V_\alpha, \sigma_\beta^\alpha)$ *and two inverse spectra* $(\overline{U}_\alpha, \bar{\rho}_\alpha^\beta)$ *and* $(\overline{V}_\alpha, \bar{\sigma}_\alpha^\beta)$ *be given* [*the elements* U_α *and* V_α ($\overline{U}_\alpha$ *and* $\overline{V}_\alpha$) *of these spectra correspond to each other in* (1-1) *manner*]. *For every* α *let a homomorphism*

$$f_\alpha \text{ of } U_\alpha \text{ into } V_\alpha; \qquad \bar{f}_\alpha \text{ of } \overline{V}_\alpha \text{ into } \overline{U}_\alpha$$

be given in such a way that for every $\beta > \alpha$ *and* $u_\alpha \in U_\alpha$ ($\bar{v}_\alpha \in \overline{V}_\alpha$) *we have*

$$\sigma_\beta^\alpha f_\alpha u_\alpha = f_\beta \rho_\beta^\alpha u_\alpha; \qquad \bar{\rho}_\alpha^\beta \bar{f}_\beta \bar{v}_\beta = \bar{f}_\alpha \bar{\sigma}_\alpha^\beta \bar{v}_\beta. \tag{23.31}$$

Then we obtain a homomorphism

$$f \text{ of } U = \limsup (U_\alpha, \rho_\alpha^\beta) \text{ into } V = \limsup (V_\alpha, \sigma_\beta^\alpha);$$

$$\bar{f} \text{ of } \overline{V} = \liminf (\overline{V}_\alpha, \bar{\sigma}_\alpha^\beta) \text{ into } \overline{U} = \liminf (\overline{U}_\alpha, \bar{\rho}_\alpha^\beta)$$

by putting in correspondence to each

bundle $u = \{u_\alpha\}$	*thread* $\bar{v} = \{\bar{v}_\alpha\}$
the bundle $v = fu$ *containing for a* $u_\alpha \in u$ *the element* $f_\alpha u_\alpha$.	*the thread* $u = \bar{f}\bar{v} = \{\bar{f}_\alpha \bar{v}_\alpha\}$.

The proof of Lemma 23.3 may be left to the reader as well as the proof of the following

LEMMA 23.4. *Under the hypotheses of Lemma* 23.3

$$\sigma_\beta^\alpha (f_\alpha U_\alpha) \subseteq f_\beta U_\beta; \qquad \bar{\rho}_\alpha^\beta (\bar{f}_\beta \overline{V}_\beta) \subseteq \bar{f}_\alpha \overline{V}_\alpha \tag{23.41}$$

and (*taking into account Lemma* 23.1)

$$\limsup (f_\alpha U_\alpha, \sigma_\beta^\alpha) = fU; \qquad \liminf (\bar{f}_\alpha \overline{V}_\alpha, \bar{\rho}_\alpha^\beta) = \bar{f}\overline{V}. \tag{23.42}$$

The same is true for the kernels of the homomorphisms $f_\alpha, f, \bar{f}_\alpha, \bar{f}$:

$$\rho_\beta^\alpha (f^{-1} 0_\alpha) \subseteq f_\beta^{-1} 0_\beta; \qquad \bar{\sigma}_\alpha^\beta (\bar{f}_\beta^{-1} \bar{0}_\beta) \subseteq \bar{f}_\alpha^{-1} \bar{0}_\alpha \tag{23.43}$$

(we write $0_\alpha, 0_\beta, \bar{0}_\alpha, \bar{0}_\beta$ for the zero elements of $V_\alpha, V_\beta, \overline{U}_\alpha, \overline{U}_\beta$), and

$$\limsup (f_\alpha^{-1} 0_\alpha, \rho_\beta^\alpha) = f^{-1} 0_V; \qquad \liminf (\bar{f}_\alpha^{-1} \bar{0}_\alpha, \bar{\sigma}_\alpha^\beta) = \bar{f}^{-1} 0_{\overline{U}}. \tag{23.44}$$

Finally we have

LEMMA 23.5. *If, still under the hypotheses of* 23.3, *we have for every* α

$$f_\alpha \left\downarrow \begin{matrix} U_\alpha & \overline{U}_\alpha \\ V_\alpha & \overline{V}_\alpha \end{matrix} \right\uparrow \bar{f}_\alpha,$$

then

$$f\begin{vmatrix} U & \overline{U} \\ \downarrow V & \overline{V} \end{vmatrix}\bar{f}.$$

The proof follows from the definition of the scalar product for limit groups [13, p. 670, formula (6.2)], namely:

For any $u=\{u_\alpha\}\in \lim\sup(U_\alpha, \rho_\beta^\alpha)$, $\bar{u}=\{\bar{u}_\alpha\}\in \lim\inf(\overline{U}_\alpha, \bar{\rho}_\alpha^\beta)$

$$(u\cdot\bar{u}) = (u_\alpha\cdot\bar{u}_\alpha), \tag{23.51}$$

where u_α *is chosen arbitrarily in* u.

The formulae (23.51) and (23.41) lead to

$$(fu\cdot\bar{v}) = (f_\alpha u_\alpha\cdot\bar{v}_\alpha) = (u_\alpha\cdot\bar{f}_\alpha\bar{v}_\alpha) = (u\cdot\bar{f}\bar{v}),$$

which proves Lemma 23.5.

24. **Simplified definitions of the ∇-groups of a cell spectrum.** Lemma 23.2 allows the definition of the group ∇_K^r, where K is a cell spectrum $K=\{K_\alpha, \varpi_\alpha^\beta\}$, to take the following simple form. Two ∇-cycles z_α^r and z_β^r of K_α, K_β, respectively, belong by definition to the same ∇-*class* of K if there exists a $\gamma>\alpha, \beta$ such that $\pi_\gamma^\alpha z_\alpha^r-\pi_\gamma^\beta z_\beta^r$ bounds on K_γ. These ∇-classes are the elements of ∇_K^r. The sum of two ∇-classes z^r and z'^r is defined as follows: take $z_\alpha^r\in z^r$, $z_\beta''\in z''$ and $\gamma>\alpha, \beta$. Then $z^r+z'^r$ is defined as the ∇-class containing $\pi_\gamma^\alpha z_\alpha^r+\pi_\gamma^\beta z_\beta^r$.

25. **The extension and the intersection homomorphisms. The groups of the figure K, A, G.** By the formulae (21.5) and (21.6) and the Lemmas 23.3 and 23.5 the homomorphisms $E_\alpha^{\alpha 0}$, $E_\alpha^{\alpha 1}$, $J_{\alpha 0}^\alpha$, $J_{\alpha 1}^\alpha$ generate the homomorphisms E_K^A, E_K^G, J_A^K, J_G^K of the corresponding limit groups according to the figure:

$$J_A^K\begin{vmatrix} \nabla_K^r & \Delta_K^r \\ \downarrow \nabla_A^r & \Delta_A^r \end{vmatrix}\uparrow E_K^A; \qquad J_G^K\begin{vmatrix} \Delta_K^r & \nabla_K^r \\ \downarrow \Delta_G^r & \nabla_G^r \end{vmatrix}\uparrow E_K^G. \tag{25.1}$$

According to Lemma 23.4 (and Lemma 23.2 used in (25.3)) define:

$$\begin{cases} \nabla_{AK}^r = J_A^K\nabla_K^r = \lim\sup(\nabla_{A_\alpha K_\alpha}^r, \pi_{\beta 0}^{\alpha 0}); \\ \Delta_{KA}^r = E_K^A\Delta_A^r = \lim\inf(\Delta_{K_\alpha A_\alpha}^r, \varpi_\alpha^\beta), \\ \Delta_{GK}^r = J_G^K\Delta_K^r = \lim\inf(\Delta_{G_\alpha K_\alpha}^r, \varpi_{\alpha 1}^{\beta 1}); \\ \nabla_{KG}^r = E_K^G\nabla_G^r = \lim\sup(\nabla_{K_\alpha G_\alpha}^r, \pi_\beta^\alpha), \end{cases} \tag{25.2}$$

$$\begin{cases} \Delta_{A:K}^r = (E_K^A)^{-1}0_K = \lim\inf(\Delta_{A_\alpha:K_\alpha}^r, \varpi_{\alpha 0}^{\beta 0}); \\ \nabla_{G:K}^r = (E_K^G)^{-1}0_K = \lim\sup(\nabla_{G_\alpha:K_\alpha}^r, \pi_{\beta 1}^{\alpha 1}), \\ \nabla_{A:K}^r = \nabla_A^r - \nabla_{AK}^r = \lim\sup(\Delta_{A_\alpha:K_\alpha}^r, \pi_{\beta 0}^{\alpha 0}); \\ \Delta_{G:K}^r = \Delta_G^r - \Delta_{GK}^r = \lim\inf(\Delta_{G_\alpha:K_\alpha}^r, \varpi_{\alpha 1}^{\beta 1}). \end{cases} \tag{25.3}$$

Then by Lemma 23.4 and the formula (12.4):

$$(J_A^K)^{-1}0_A = \limsup (\nabla^r_{K_\alpha G_\alpha}, \pi^\alpha_\beta) = \nabla^r_{KG};$$

$$(J_G^K)^{-1}0_G = \liminf (\Delta^r_{K_\alpha A_\alpha}, \varpi^\beta_\alpha) = \Delta^r_{KA}. \tag{25.4}$$

From these formulae we get literally as in section 14 the relations (14.1)–(14.2) for the spectra K, A, G.

26. **A direct definition of the groups of the figure** K, A, G. Since any Δ-cycle (Δ-homology) on A_α is a Δ-cycle (a Δ-homology) on K_α, to any thread $z^r_A = \{\mathfrak{Z}^r_{\alpha 0}\} \in \Delta^r_A$ corresponds the thread $z^r_K = \{\mathfrak{Z}^r_\alpha\} \in \Delta^r_K$ where $\mathfrak{Z}^r_\alpha$ is completely determined by $\mathfrak{Z}^r_{\alpha 0} \subseteq \mathfrak{Z}^r_\alpha$. This thread z^r_K is the thread $E^A_K z^r_A$.

Since any ∇-cycle (∇-homology) on G_α is a ∇-cycle (∇-homology) on K_α, any ∇-class (section 24) $z^r_G \in \nabla^r_G$ is a subset of a well determined ∇-class $z^r_K \in \nabla^r_K$, and this z^r_K is the ∇-class $E^G_K z^r_G$.

Let $z^r_K = \{\mathfrak{Z}^r_\alpha\} \in \Delta^r_K$. Then $\{J^\alpha_{\alpha 1}\mathfrak{Z}^r_\alpha\}$ is a thread and this thread is the thread $J^K_G z^r_K \in \Delta^r_G$.

Let $z^r_K = \{z^r_\alpha\} \in \nabla^r_K$. Then all $J^\alpha_{\alpha 0} z^r_\alpha$ are elements of the same ∇-class $z^r_A \in \nabla^r_A$ and this ∇-class is $J^K_A z^r_K$.

Thus we get the following direct definitions of the groups of the figure K, A, G:

The elements of $\Delta^r_{KA} \subseteq \Delta^r_K$ are threads $z^r_K = \{\mathfrak{Z}^r_\alpha\} \in \Delta^r_K$ containing threads $z^r_A = \{\mathfrak{Z}^r_{\alpha 0}\} \in \Delta^r_A$ in the sense that each Δ-class $\mathfrak{Z}^r_{\alpha 0}$ is a subset of the corresponding Δ-class $\mathfrak{Z}^r_\alpha$.

The elements of ∇^r_{KG} are ∇-classes $z^r_K \in \nabla^r_K$ containing as subsets ∇-classes $z^r_G \in \nabla^r_G$.

The elements of $\nabla^r_{AK} \subseteq \nabla^r_A$ are ∇-classes $z^r_A \in \nabla^r_A$ containing among their elements extensible ∇-cycles, that is, ∇-cycles $z^r_{\alpha 0}$ on A_α of the form $z^r_{\alpha 0} = J^\alpha_{\alpha 0} z^r_\alpha$ where z^r_α is a ∇-cycle on K_α. For any element $z^r_{\alpha 0}$ of $z^r_A \in \nabla^r_{AK}$ there is a $\beta > \alpha$ such that $\pi^{\alpha 0}_{\beta 0} z^r_{\alpha 0}$ is extensible.

The elements of $\Delta^r_{GK} \subseteq \Delta^r_G$ are "extensible threads" $z^r_G = \{\mathfrak{z}^r_{\alpha 1}\} \in \Delta^r_G$, that is, threads whose elements $\mathfrak{z}^r_{\alpha 1}$ are all extensible (that is of the form $\mathfrak{z}^r_{\alpha 1} = J^\alpha_{\alpha 1}\mathfrak{z}^r_\alpha$ where $\mathfrak{z}^r \in \Delta^r_\alpha$.

The elements of $\Delta^r_{A:K} \subseteq \Delta^r_A$ are threads $z^r_A = \{\mathfrak{z}^r_{\alpha 0}\} \in \Delta^r_A$ *bounding on* K in the sense that the elements of each Δ-class $\mathfrak{z}^r_{\alpha 0}$ bound on K_α.

The elements of $\nabla^r_{G:K} \subseteq \nabla^r_G$ are ∇-classes $z^r_G \in \nabla^r_G$ *bounding on* K in the sense that z^r_G is a subset of the ∇-class $z^r_K = 0$ representing the zero of the group ∇^r_K. In other words, for any $z^r_{\alpha 1} \in z^r_G$ there is a $\pi^{\alpha 1}_{\beta 1} z^r_{\alpha 1}$ bounding on K_β.

DEFINITION 26.1. *The ∇-cycle $z^r_{\alpha 0}$ on A_α and the ∇-cycle $z^r_{\beta 0}$ on A_β belong to the same extension class (with respect to the cell spectrum K) if there exists a $\gamma > \alpha, \beta$ such that $\pi^{\alpha 0}_{\gamma 0} z^r_{\alpha 0} - \pi^{\beta 0}_{\gamma 0} z^r_{\beta 0}$ is extensible.*

DEFINITION 26.2. *The threads $z^r_G = \{\mathfrak{Z}^r_{\alpha 1}\}$ and $z^{r'}_G = \{\mathfrak{Z}^{r'}_{\alpha 1}\}$ belong to the same extension class if for each α the Δ-class $\mathfrak{z}^r_{\alpha 1} - \mathfrak{z}^{r'}_{\alpha 1}$ is extensible (that is, is of the form $\mathfrak{z}^r_{\alpha 1} - \mathfrak{z}^{r'}_{\alpha 1} = J^\alpha_{\alpha 1}\mathfrak{z}^r_\alpha$ with $z^r_\alpha \in \Delta^r_\alpha$).*

Now the elements of $\nabla'_{A:K}=\nabla'_A-\nabla'_{AK}$ can be identified with the extension classes of ∇-cycles $z'_{\alpha 0}$. To get the sum of two extension classes z^r and z'^r take as always $z'_{\alpha 0}\in z^r$, $z''_{\beta 0}\in z'^r$, choose any $\gamma>\alpha$, β and define $z^r+z'^r$ as the extension class containing $\pi^{\alpha 0}_{\gamma 0}z'_{\alpha 0}+\pi^{\beta 0}_{\gamma 0}z''_{\beta 0}$; the group $\nabla'_{A:K}$ can be defined as the group of the extension classes of r-dimensional ∇-cycles on A_α (α variable) with this addition.

The elements of the group $\Delta'_{G:K}$ are extension classes of threads $z'_G\in\Delta'_G$.

27. **Proof of the third pair of isomorphisms.** Let

$$z^r_A=\{z^r_{\alpha 0}\}\in\nabla^r_{A:K}.$$

The ∇-cycle $\nabla E^{\alpha 0}_\alpha z'_{\alpha 0}$ is on G_α and bounds on K_α.

27.1. If $z'_{\alpha 0}$ and $z'_{\beta 0}$ belong to the same extension class, that is, if for a certain $\gamma>\alpha,\beta$

$$\pi^{\alpha 0}_{\gamma 0}z^r_{\alpha 0}-\pi^{\beta 0}_{\gamma 0}z^r_{\beta 0}=J^\gamma_{\gamma 0}z^r_\gamma, \tag{27.11}$$

where z'_γ is a ∇-cycle on K_γ, then

$$\pi^\alpha_\gamma\nabla E^{\alpha 0}_\alpha z^r_{\alpha 0}-\pi^\beta_\gamma\nabla E^{\beta 0}_\beta z^r_{\beta 0}\in H^{r+1}_{\gamma 1} \tag{27.12}$$

($H^r_{\gamma 1}$ is the group of all r-dimensional ∇-cycles on G_γ bounding on G_γ).

To prove (27.12) we first deduce from (27.11) and section 15:

$$\nabla E^{\gamma 0}_\gamma\pi^{\alpha 0}_{\gamma 0}z^r_{\alpha 0}-\nabla E^{\gamma 0}_\gamma\pi^{\beta 0}_{\gamma 0}z^r_{\beta 0}\in H^{r+1}_{\gamma 1}. \tag{27.13}$$

On the other hand

$$\pi^\alpha_\gamma\nabla E^{\alpha 0}_\alpha z^r_{\alpha 0}-\nabla E^{\gamma 0}_\gamma\pi^{\alpha 0}_{\gamma 0}z^r_{\alpha 0}=\nabla(\pi^\alpha_\gamma E^{\alpha 0}_\alpha z^r_{\alpha 0}-E^{\gamma 0}_\gamma\pi^{\alpha 0}_{\beta 0}z^r_{\alpha 0})$$

and $\pi^\alpha_\gamma E^{\alpha 0}_\alpha z'_{\alpha 0}-E^{\gamma 0}_\gamma\pi^{\alpha 0}_{\gamma 0}z'_{\alpha 0}$ is on G_γ. Thus

$$\pi^\alpha_\gamma\nabla E^{\alpha 0}_\alpha z^r_{\alpha 0}-\nabla E^{\gamma 0}_\gamma\pi^{\alpha 0}_{\gamma 0}z^r_{\alpha 0}\in H^{r+1}_{\gamma 1} \tag{27.14}$$

and in the same way

$$\pi^\beta_\gamma\nabla E^{\beta 0}_\gamma z^r_{\beta 0}-\nabla E^{\gamma 0}_\gamma\pi^{\beta 0}_{\gamma 0}z^r_{\beta 0}\in H^{r+1}_{\gamma 1}. \tag{27.15}$$

From (27.13)–(27.15) follows (27.12) and therefore (27.1).

Thus the operators $\nabla E^{\alpha 0}_\alpha$ generate a homomorphism—we call it ∇E—of $\nabla'_{A:K}$ into $\nabla'^{+1}_{G:K}$.

27.2. *The homomorphism* ∇E *maps* $\nabla'_{A:K}$ *onto* $\nabla'^{+1}_{G:K}$. To prove this let $z'^{+1}_G\in\nabla'^{+1}_{G:K}$ and $z'^{+1}_{\alpha 1}\in z'^{+1}_G$. We are seeking a $\beta>\alpha$ and a ∇-cycle $z'_{\beta 0}$ on A_β such that

$$\nabla E^{\beta 0}_\beta z^r_{\beta 0}-\pi^{\alpha 1}_{\beta 1}z^{r+1}_{\alpha 1}\in H^{r+1}_{\beta 1}. \tag{27.21}$$

From $z'^{+1}_{\alpha 1}\in z'^{+1}_G\in\nabla'^{+1}_{G:K}$ and the definition of $\nabla'^{+1}_{G:K}$ it follows that there exist a $\beta>\alpha$ and a chain x'_β on K_β such that $\nabla x'_\beta=\pi^{\alpha 1}_{\beta 1}z'^{+1}_{\alpha 1}$. As $\nabla x'_\beta$ is on G_β, the chain

$J^{\beta}_{\beta 0}x^r_\beta = z^r_{\beta 0}$ is a ∇-cycle on A_β, while $x^r_\beta - E^{\beta 0}_\beta J^{\beta}_{\beta 0}x^r_\beta$ is on G and

$$\nabla(x^r_\beta - E^{\beta 0}_\beta J^{\beta}_{\beta 0}x^r_\beta) = \nabla x^r_\beta - \nabla E^{\beta 0}_\beta J^{\beta}_{\beta 0}x^r_\beta = \pi^{\alpha 1}_{\beta 1} z^{r+1}_{\alpha 1} - \nabla E^{\beta 0}_\beta z^r_{\beta 0}$$

from which follows (27.21) and thus (27.2).

27.3. *∇E maps $\nabla^r_{A:K}$ on $\nabla^{r+1}_{G:K}$ isomorphically.* Suppose that the element $z^r_A \in \nabla^r_{A:K}$ is mapped by ∇E on the zero element of the group $\nabla^{r+1}_{G:K}$. We shall prove that the extension class z^r_A is the zero element of the group $\nabla^r_{A:K}$. Take $z^r_{\alpha 0} \in z^r_A$. It follows from our hypotheses that there exists such a $\beta > \alpha$ that

$$\pi^{\alpha}_{\beta} \nabla E^{\alpha 0}_{\alpha} z^r_{\alpha 0} \in H^{r+1}_{\beta 1}.$$

Since

$$\pi^{\alpha}_{\beta} \nabla E^{\alpha 0}_{\alpha} z^r_{\alpha 0} - \nabla E^{\beta 0}_{\beta} \pi^{\alpha 0}_{\beta 0} z^r_{\alpha 0} \in H^{r+1}_{\beta 1}$$

(see (27.14)) and $\pi^{\alpha 0}_{\beta 0} z^r_{\alpha 0} \in z^r_A$, we may suppose from the beginning (replacing β by α and $\pi^{\alpha 0}_{\beta 0} z^r_{\alpha 0}$ by $z^r_{\alpha 0}$) that for some $z^r_{\alpha 0} \in z^r_A$

$$\nabla E^{\alpha 0}_{\alpha} z^r_{\alpha 0} \in H^{r+1}_{\alpha 1}.$$

To deduce from that the identity $z^r_A = 0$ we have only to prove that $z^r_{\alpha 0}$ is extensible. To this end take a chain $x^r_{\alpha 1}$ on G_α bounded by $\nabla E^{\alpha 0}_{\alpha} z^r_{\alpha 0}$. Then $z^r_\alpha = E^{\alpha 0}_{\alpha} z^r_{\alpha 0} - x^r_{\alpha 1}$ is a ∇-cycle on K_α and $J^{\alpha}_{\alpha 0} z^r_\alpha = z^r_{\alpha 0}$, q.e.d.

§7. Simplicial spectra

28. **Nerves.** We denote by N, N_α, N_β, $\cdots$ and call *nerves* the finite complete simplicial complexes (complete means every face of an element of N is itself an element of N). The oriented simplices of N form a cell complex denoted by N too.

Let N_α, N_β be nerves, C_α, C_β closed subcomplexes of N_α, N_β, respectively, and

$$K_\alpha = N_\alpha \backslash C_\alpha, \qquad K_\beta = N_\beta \backslash C_\beta.$$

Let σ^{β}_{α} be a simplicial mapping of N_β into N_α such that the image of C_β under σ^{β}_{α} is contained in C_α:

$$\sigma^{\beta}_{\alpha} C_\beta \subseteq C_\alpha. \tag{28}$$

Then in a well known way σ^{β}_{α} generates a projection denoted also by σ^{β}_{α} of the cell complex N_β into the cell complex N_α; for any oriented simplex $t^r_\beta = (e^{\beta}_0 \cdots e^{\beta}_r)$ of N_β we put $\sigma^{\beta}_{\alpha} t^r_\beta = (\sigma^{\beta}_{\alpha} e^{\beta}_0 \cdots \sigma^{\beta}_{\alpha} e^{\beta}_r)$ if the vertices $\sigma^{\beta}_{\alpha} e^{\beta}_0, \cdots, \sigma^{\beta}_{\alpha} e^{\beta}_r$ are all different, and $\sigma^{\beta}_{\alpha} t^r_\beta = 0$ if for some $i \neq j$, $\sigma^{\beta}_{\alpha} e^{\beta}_i = \sigma^{\beta}_{\alpha} e^{\beta}_j$, whereupon for any chain $x^r_\beta = \sum a_i t^r_{\beta i}$, $\sigma^{\beta}_{\alpha} x^r_\beta = \sum a_i \sigma^{\beta}_{\alpha} t^r_{\beta i}$. The projection σ^{β}_{α} of N_β into N_α generates a projection ϖ^{β}_{α} of K_β into K_α by setting

$$\varpi^{\beta}_{\alpha} x^r = J^{N_\alpha}_{K_\alpha} \sigma^{\beta}_{\alpha} x^r_\beta \qquad \text{for any } x^r_\beta \in L^r_\beta = L^r_{K_\beta}$$

(see section 20 where K_α, G_α, ϖ^{β}_{α}, $\varpi^{\beta 1}_{\alpha 1}$ are to be replaced by N_α, K_α, σ^{β}_{α}, ϖ^{β}_{α}, respectively).

A known fact is(4):

LEMMA 28.1. *Let σ_α^β and $\sigma_\alpha'^\beta$ be two simplicial mappings of N_β into N_α both satisfying the condition* (28) *and the condition*:

(28.11) *For every simplex $T_\beta \in N_\beta$ there exists a simplex $T_\alpha \in N_\alpha$ containing among its vertices all vertices of both $\sigma_\alpha^\beta T_\beta$ and $\sigma_\alpha'^\beta T_\beta$; if in particular $T_\beta \in C_\beta$ then T_α can be chosen among the simplices of C_α.*

Under these hypotheses the projections ϖ_α^β and $\varpi_\alpha'^\beta$ of K_β into K_α generated by σ_α^β and $\sigma_\alpha'^\beta$ are homologous to each other.

The proof is as follows. Let $e_1, e_2, \cdots, e_s$ be all vertices of N_β in a definite order. Take a copy N_β' of N_β with the vertices $e_1', e_2', \cdots, e_s'$ and consider all simplices of the form $|e_{i_0}' \cdots e_{i_p}' e_{i_p} \cdots e_{i_r}|$, where $p=0, 1, \cdots, r$ and $|e_{i_0} \cdots e_{i_r}|$ is any simplex of N_β. The simplices just defined and their faces form a nerve N_β'' called the *prism* over N_β. Now define for each oriented simplex $t_\beta^r=(e_{i_0} \cdots e_{i_r})$ of N_β the *prism over* t_β^r as the chain

$$\Pi t_\beta^r = \sum_{p=0}^{p=r} (-1)^p (e_{i_0}' \cdots e_{i_p}' e_{i_p} \cdots e_{i_r})$$

while the prism over the chain $x_\beta^r = \sum a_i t_{\beta i}^r \in L_{N_\beta}^r$ is by definition the chain

$$\Pi x_\beta^r = \sum a_i \Pi t_{\beta i}^r$$

on N_β''. An easy calculation shows: *If z_β^r is a Δ-cycle on K_β then*

$$\Delta \Pi z_\beta^r = z_\beta^r - z_\beta'^r - \Pi \Delta z_\beta^r, \tag{28.12}$$

where $z_\beta''^r$ is the copy of z_β^r on N_β' and Δ operates on N_β'' at the left and on N_β at the right (see [4, p. 199]).

Now map the vertices of N_β by σ_α^β and those of N_β' by $\sigma_\alpha'^\beta$. By virtue of (28.11) we get in this way a simplicial mapping $\sigma_\alpha''^\beta$ of N_β'' into N_α. Denoting by x_α^{r+1} the chain $\sigma_\alpha''^\beta \Pi z_\beta^r$ (on N_α), by x_α^r the chain $\sigma_\alpha''^\beta \Pi \Delta z_\beta^r$ which is on C_α, we get

$$\Delta_{N_\alpha} x_\alpha^{r+1} = \sigma_\alpha^\beta z_\beta^r - \sigma_\alpha'^\beta z_\beta^r - x_\alpha^r$$

and, K_α being open on N_α,

$$\Delta_{K_\alpha} J_{K_\alpha}^{N_\alpha} x_\alpha^{r+1} = J_{K_\alpha}^{N_\alpha} \Delta_{N_\alpha} x_\alpha^{r+1} = J_{K_\alpha}^{N_\alpha} \sigma_\alpha^\beta z_\beta^r - J_{K_\alpha}^{N_\alpha} \sigma_\alpha'^\beta z_\beta^r = \varpi_\alpha^\beta z_\beta^r - \varpi_\alpha'^\beta z_\beta^r$$

q.e.d.

In the next chapter (section 31) we need the following

COROLLARY 28.2. *Let A_α, A_β be closed subcomplexes of $K_\alpha = N_\alpha \backslash C_\alpha$, $K_\beta = N_\beta \backslash C_\beta$; let σ_α^β, $\sigma_\alpha'^\beta$ map C_β into C_α as well as A_β into A_α. Suppose the con-*

(4) We write $T^r = |e_0 \cdots e_r|$ for non-oriented and $t^r = (e_0 \cdots e_r)$ for oriented simplices.

dition (28.11) *satisfied; suppose moreover that for* $T_\beta \in A_\beta$ *the simplex* T_α *of the latter condition can always be chosen in* A_α. *Under these hypotheses not only the projections* ϖ_α^β *and* $\varpi_\alpha'^\beta$ *of* K_β *into* K_α *generated by* σ_α^β, $\sigma_\alpha'^\beta$ *are homologous but the same is true for the corresponding projections* $\varpi_{\alpha 0}^{\beta 0}$, $\varpi_{\alpha 0}'^{\beta 0}$ *and* $\varpi_{\alpha 1}^{\beta 1}$, $\varpi_{\alpha 1}'^{\beta 1}$ *of* A_β *into* A_α *and of* G_β *into* G_α.

To prove the assertion concerning A_β, A_α replace in 28.1 N_β, N_α by $A_\beta \cup C_\beta$, $A_\alpha \cup C_\alpha$ and K_β, K_α by A_β, A_α. To prove the assertion concerning G_β, G_α replace in 28.1 C_β, C_α by $C_\beta \cup A_\beta$, $C_\alpha \cup A_\alpha$ and K_β, K_α by G_β, G_α.

29. **Simplicial spectra.** Let an unbounded partially ordered set of nerves N_α be given; we write $\beta > \alpha$ if in this set N_β follows on N_α. Let (for each α) C_α be a fixed closed subcomplex of N_α. We suppose that for each pair $\beta > \alpha$ a finite number of simplicial mappings σ_α^β (called *projections*) of N_β into N_α is given and that the following conditions are satisfied:

1. σ_α^β maps C_β into C_α;
2. for each pair $\beta > \alpha$ the projections σ_α^β satisfy the conditions (28.11);
3. if $\gamma > \beta > \alpha$ and σ_β^γ, σ_α^β are projections, then the simplicial mapping $\sigma_\alpha^\gamma = \sigma_\alpha^\beta \sigma_\beta^\gamma$ of N_γ into N_α is a projection.

Under these hypotheses the partially ordered set of nerves N_α, of their subcomplexes C_α and of the projections σ_α^β is called a *simplicial spectrum*

$$N = \{N_\alpha, C_\alpha, \sigma_\alpha^\beta\}. \tag{29.1}$$

A simplicial spectrum (29.1) generates according to section 28 the *cell spectrum*

$$K = \{K_\alpha, \varpi_\alpha^\beta\}, \tag{29.2}$$

where by K_α is denoted the open subcomplex $N_\alpha \setminus C_\alpha$ as well as the cell complex of all oriented simplices of this open subcomplex; the projections ϖ_α^β are (as defined in section 28):

$$\varpi_\alpha^\beta x_\beta^r = J_{K_\alpha}^{N_\alpha} \sigma_\alpha^\beta x_\beta^r \qquad \text{for any } x_\beta^r \in L_\beta^r = L_{K_\beta}^r.$$

REMARK. The complexes C_α are called *special subcomplexes* of the spectrum N; if they vanish the spectrum is called *bicompact.*

CHAPTER III. SPACES

§8. THE CELL SPECTRUM AND THE BETTI GROUPS OF A LOCALLY BICOMPACT SPACE

30. **The simplicial and the cell spectrum of a locally bicompact space.** The nerve N is called the *nerve of the finite system*

$$\alpha = \{A_1, \cdots, A_s\}$$

of subsets A_i (of the given set R) if the vertices $e_1, \cdots, e_s$ of N are in (1-1)-

correspondence with the elements $A_1, \cdots, A_s$ of the system α and any $e_{i_0}, \cdots, e_{i_r}$ are vertices of a simplex of N if and only if the *associate sets* $A_{i_0}, \cdots, A_{i_r}$ have a non-vacuous intersection(5).

Now let $\mathbf{K}$ be a locally bicompact Hausdorff space. We consider the set of all finite coverings,

$$\alpha = \{o_{\alpha 1}, \cdots, o_{\alpha s}\}, \qquad s = s_\alpha,$$

of $\mathbf{K}$ by open sets $o_{\alpha 1}, \cdots, o_{\alpha s}$. The nerve of α is denoted by N_α. We denote by C_α the subcomplex of N_α consisting of all simplices of N_α whose vertices correspond to sets $o_{\alpha i} \in \alpha$ having a non-bicompact closure $\delta_{\alpha i}$ in $\mathbf{K}$. Thus the open subcomplex $K_\alpha = N_\alpha \setminus C_\alpha$ is the subcomplex of all simplices of N_α which have among their vertices at least one corresponding to a set $o_{\alpha i} \in \alpha$ with bicompact closure. The set of all N_α is partially ordered: we set $\beta > \alpha$ if β is a refinement of α (that is if each element of β is a subset of one or more elements of α) while α is not a refinement of β.

Let $\beta > \alpha$. We get by definition a projection σ_α^β of N_β into N_α if we assign to each vertex $e_{\beta j}$ of N_β any vertex $e_{\alpha i}$ of N_α satisfying the condition that $o_{\beta j} \subseteq o_{\alpha i}$. It is easily seen that the nerves N_α with their special subcomplexes C_α and projections σ_α^β constitute a simplicial spectrum $\{N_\alpha, C_\alpha, \sigma_\alpha^\beta\}$ called *the simplicial spectrum of the locally bicompact space* $\mathbf{K}$. According to section 29 this simplicial spectrum generates the cell spectrum

$$K = \{K_\alpha, \varpi_\alpha^\beta\} \tag{30}$$

called *the cell spectrum of the locally bicompact space* $\mathbf{K}$. The Betti groups

$$\Delta_K^r = \liminf (\Delta_{K_\alpha}^r, \varpi_\alpha^\beta); \qquad \nabla_K^r = \limsup (\nabla_{K_\alpha}^r, \pi_\beta^\alpha)$$

of the cell spectrum K are called *Betti groups of the space* $\mathbf{K}$ and are denoted correspondingly by $\Delta_{\mathbf{K}}^r, \nabla_{\mathbf{K}}^r$.

If the space $\mathbf{K}$ is bicompact, then so is its spectrum K and the previous definitions are simplified by $C_\alpha = 0$, $K_\alpha = N_\alpha$.

31. **The figures $\mathbf{K}$, $\mathbf{A}$, $\mathbf{\Gamma}$ and K, A, G.** Let the notations of the previous section hold. Let $\mathbf{A}$ be a closed set in the locally bicompact *normal* space $\mathbf{K}$; the complement $\mathbf{\Gamma} = \mathbf{K} \setminus \mathbf{A}$ is open.

Denote by $A_\alpha \subseteq K_\alpha$ the closed subcomplex of all simplices $T_\alpha \in K_\alpha$ satisfying the condition: *the associate set $o_{\alpha i} \in \alpha$ of each vertex $e_{\alpha i}$ of T_α has points in common with* $\mathbf{A}$. It is easy to see that the projections σ_α^β of N_β into N_α defined in the last section satisfy the conditions of the Corollary 28.2 and thus define in addition to the cell spectrum (30) the cell spectra

$$A = \{A_\alpha, \varpi_{\alpha 0}^{\beta 0}\}, \tag{31.1}$$

(5) Our terminology is correct, each nerve being, for example, the nerve of the system of the open stars of its own vertices.

and

$$G = \{G_\alpha, \varpi_{\alpha 1}^{\beta 1}\}. \tag{31.2}$$

Now $\Delta_{\mathbf{K}}^r$, $\nabla_{\mathbf{K}}^r$ are by definition identical with Δ_K^r, ∇_K^r while, as we shall see in the next paragraph, there exist certain natural isomorphisms between the groups

$$\Delta_A^r \text{ and } \Delta_{\mathbf{A}}^r; \qquad \nabla_A^r \text{ and } \nabla_{\mathbf{A}}^r,$$

$$\Delta_G^r \text{ and } \Delta_{\mathbf{\Gamma}}^r; \qquad \nabla_G^r \text{ and } \nabla_{\mathbf{\Gamma}}^r.$$

These isomorphisms transform the homomorphisms E_K^A, E_K^G, J_A^K, J_G^K into the homomorphisms $E_{\mathbf{K}}^{\mathbf{A}}$, $E_{\mathbf{K}}^{\mathbf{\Gamma}}$, $J_{\mathbf{A}}^{\mathbf{K}}$, $J_{\mathbf{\Gamma}}^{\mathbf{K}}$ and thus lead to the definition of the groups of the figure $\mathbf{K}$, $\mathbf{A}$, $\mathbf{\Gamma}$.

§9. The natural isomorphisms, the groups and the homomorphisms of the figure $\mathbf{K}$, $\mathbf{A}$, $\mathbf{\Gamma}$

32. Special types of coverings. Throughout this chapter $\mathbf{K}$ means always a normal locally bicompact space; by a *covering* (without any special attribute) is always meant a finite covering by open sets (of the given space).

In any covering $\alpha = \{o_{\alpha 1}, \cdots, o_{\alpha s}\}$ of $\mathbf{K}$ we denote by $o_{\alpha 1}, \cdots, o_{\alpha p}$ and call *elements of the first kind* the elements meeting $\mathbf{A}$; the remaining elements $o_{\alpha,p+1}, \cdots, o_{\alpha s}$ have no points in common with $\mathbf{A}$ and are called *elements of the second kind*. An element of the second kind is called a *boundary* or an *inner* element according to whether its closure has or has not points in common with $\mathbf{A}$.

A covering $\alpha = \{o_{\alpha 1}, \cdots, o_{\alpha p}, \cdots, o_{\alpha s}\}$ of $\mathbf{K}$ is called *cogredient* with $\mathbf{A}$ if it satisfies the following conditions:

1. Any two among the point sets $\mathbf{A} \cap o_i$, $i \leqq p$, are different.
2. If $i_0, \cdots, i_r \leqq p$ and $o_{\alpha i_0} \cap \cdots \cap o_{\alpha i_r} \neq 0$, then $\mathbf{A} \cap o_{\alpha i_0} \cap \cdots \cap o_{\alpha i_r} \neq 0$.
3. The bicompactness of $\overline{\mathbf{A} \cap o_{\alpha i}}$, $i \leqq p$, implies that of $\delta_{\alpha i}$.

Let $\alpha = \{o_{\alpha 1}, \cdots, o_{\alpha p}, \cdots, o_{\alpha s}\}$ be a covering of $\mathbf{K}$. Denote by ϕ the sum of all those sets $\delta_{\alpha i}$ which are bicompact and lie in $\mathbf{\Gamma}$. The covering α is called *cogredient with* $\mathbf{\Gamma}$, if it satisfies the following conditions:

1. The elements of the second kind of α form a covering of $\mathbf{\Gamma}$ (denoted by $\mathbf{\Gamma}\alpha$).
2. No element of the first kind meets ϕ.

REMARK. In my paper [1] coverings cogredient with $\mathbf{\Gamma}$ were called "regular with respect to $\mathbf{\Gamma}$."

The covering $\alpha = \{o_{\alpha 1}, \cdots, o_{\alpha p}, \cdots, o_{\alpha s}\}$ is called *regular* ("regular with respect to $\mathbf{A}$" in the terminology of [1]) if the following conditions are fulfilled:

1. The covering α contains no boundary element of the second kind.
2. If $\delta_{\alpha i_0} \cap \cdots \cap \delta_{\alpha i_r} \neq 0$ and $i_0, \cdots, i_r \leqq p$, then $\mathbf{A} \cap o_{\alpha i_0} \cap \cdots \cap o_{\alpha i_r} \neq 0$.

To finish with these preliminary notions take in K_α all simplices T_α with the following property: Each vertex of T_α corresponds either to an element $o_{\alpha i}\in\alpha$ of the first kind or to a boundary element of the second kind; these simplices form a closed subcomplex A'_α of K_α and for $\beta>\alpha$ the complex A'_β is mapped by any σ^β_α into A'_α, the conditions of Corollary 28.2 being satisfied. Therefore we have the cell spectrum $A'=\{A'_\alpha, \varpi'^{\beta 0}_{\alpha 0}\}$ and the still more important for us cell spectrum

$$G' = \{G'_\alpha, \varpi'^{\beta 1}_{\alpha 1}\}, \tag{32.1}$$

where $G'_\alpha = K_\alpha \setminus A'_\alpha$, and $\varpi'^{\beta 0}_{\alpha 0}$, $\varpi'^{\beta 1}_{\alpha 0}$ are the projections of A'_β, G'_β into A'_α, G'_α generated by σ^β_α.

32.2. Obviously if α is regular, then $A'_\alpha = A_\alpha$, $G'_\alpha = G_\alpha$.

The following results are proved in detail in my paper [1] (one can read the proofs without reading the rest of the paper):

32.3. Any covering of $\mathbf{K}$ has a regular refinement [1, p. 80, Theorem 6.22].

32.4. Any covering of $\mathbf{K}$ has a refinement cogredient with Γ—and even more:

32.5. Any two coverings α and β of $\mathbf{K}$ have such a common refinement γ cogredient with Γ that every element of the second kind of γ is contained in an element of the second kind of α and in an element of the second kind of β [1, p. 89, Lemma 8.5].

32.6. For every covering α_Γ of Γ there is a covering α of $\mathbf{K}$ cogredient with Γ and such that $\alpha_\Gamma = \Gamma\alpha$ [1, pp. 87–88, Theorem 8.33].

Let us prove now:

32.7. *Every covering* $\alpha=\{o_{\alpha 1}, \cdots, o_{\alpha p}, \cdots, o_{\alpha s}\}$ *of* $\mathbf{K}$ *has a refinement cogredient with* $\mathbf{A}$.

Proof. Let

$$\mathbf{A}\cap o_{\alpha 1}, \cdots, \mathbf{A}\cap o_{\alpha q}, \qquad q \leqq p, \tag{32.71}$$

be chosen among the $\mathbf{A}\cap o_{\alpha 1}, \cdots, \mathbf{A}\cap o_{\alpha p}$ in such a manner that no proper subsystem of the system (32.71) is a covering of $\mathbf{A}$. Take such closed sets $a_i\subseteq(\mathbf{A}\cap o_{\alpha i})$, $i=1, 2, \cdots, q$, as to get a closed covering of $\mathbf{A}$. Choose for each a_i a neighborhood $Oa_i\subseteq o_{\alpha i}$ under the following conditions:

1. All Oa_i are different.
2. Any $Oa_{i_0}, \cdots, Oa_{i_r}$ have points in common only if $a_{i_0}, \cdots, a_{i_r}$ have.
3. The bicompactness of $\bar{a}_i$ implies that of $\overline{Oa_i}$. The sum of $Oa_1, \cdots, Oa_q$ is a neighborhood $O\mathbf{A}$ of $\mathbf{A}$. Take the closed sets $b_h\subset\mathbf{K}\setminus O\mathbf{A}$ so as to get a closed covering $\{b_1, \cdots, b_u\}$ of $\mathbf{K}\setminus O\mathbf{A}$ in such a way that each of the b_h is contained in some $o_{\alpha i}$. Then define Ob_h as $o_{\alpha i}\cap(\mathbf{K}\setminus\mathbf{A})$, where $o_{\alpha i}$ is any element of α containing b_h. The covering $Oa_1, \cdots, Oa_q, Ob_1, \cdots, Ob_u$ has the required properties.

Let $\alpha_\mathbf{A}$ be a covering of $\mathbf{A}$. We write $\alpha_\mathbf{A}=\mathbf{A}\alpha$, α being a covering of $\mathbf{K}$, if the elements of $\alpha_\mathbf{A}$ are in (1-1)-correspondence with the elements of the

first kind of α, each element of $\alpha_{\mathbf{A}}$ being the intersection of $\mathbf{A}$ with the corresponding element of α. From 32.7 follows:

32.8. If we consider among the coverings of $\mathbf{A}$ only those of the form $\alpha_A = \mathbf{A}\alpha$, where α is a covering of $\mathbf{K}$ cogredient with A, and set for these coverings $\mathbf{A}\beta > \mathbf{A}\alpha$ only if $\beta > \alpha$ we get a cofinal part of the set of all coverings of $\mathbf{A}$.

33. A lemma on group spectra.

LEMMA 33. *Let the group spectra:*

$$(\textit{inverse})\ \{\overline{U}_{\alpha\lambda}, \varpi^{\beta\mu}_{\alpha\lambda}\}\ \textit{and}\ \{\overline{U}_\alpha, \varpi^\beta_\alpha\};\ (\textit{direct})\ \{U_{\alpha\lambda}, \pi^{\alpha\lambda}_{\beta\mu}\}\ \textit{and}\ \{U_\alpha, \pi^\alpha_\beta\}$$

satisfy the following conditions:

1. *There exists an isomorphism* $\phi^{\alpha\lambda}_\alpha$

$$\textit{of } \overline{U}_{\alpha\lambda} \textit{ onto } \overline{U}_\alpha; \textit{ of } U_{\alpha\lambda} \textit{ onto } U_\alpha.$$

2. $\beta\mu > \alpha\lambda$ *implies* $\beta > \alpha$.
3. *For every*

$$\bar{u}_{\beta\mu} \in \overline{U}_{\beta\mu}; \qquad u_{\alpha\lambda} \in U_{\alpha\lambda}$$

we have

$$\phi^{\alpha\lambda}_\alpha \varpi^{\beta\mu}_{\alpha\lambda} \bar{u}_{\beta\mu} = \varpi^\beta_\alpha \phi^{\beta\mu}_\beta \bar{u}_{\beta\mu}; \qquad \phi^{\beta\mu}_\beta \pi^{\alpha\lambda}_{\beta\mu} u_{\alpha\lambda} = \pi^\alpha_\beta \phi^{\alpha\lambda}_\alpha u_{\alpha\lambda}.$$

Under these hypotheses if we replace the elements of

$$\textit{each thread } \{\bar{u}_{\alpha\lambda}\};\ \textit{each bundle } \{u_{\alpha\lambda}\}$$

by their images under $\phi^{\alpha\lambda}_\alpha$ *we get an isomorphism between*

$$\liminf (\overline{U}_{\alpha\lambda}, \varpi^{\beta\mu}_{\alpha\lambda}) \textit{ and } \liminf (\overline{U}_\alpha, \varpi^\beta_\alpha);\ \limsup (U_{\alpha\lambda}, \pi^{\alpha\lambda}_{\beta\mu}) \textit{ and } \limsup (U_\alpha, \pi^\alpha_\beta).$$

The proof is easy and may be found in [1, p. 62, Theorem 3.61].

34. The natural isomorphisms between Δ^r_A, ∇^r_A and $\Delta^r_{\mathbf{A}}$, $\nabla^r_{\mathbf{A}}$. We return to the figures of section 31 but we change slightly our notations: we denote now by α any covering of $\mathbf{A}$ which can be written in the form $\alpha = \mathbf{A}\alpha\lambda$, where $\alpha\lambda$ is a covering of $\mathbf{K}$ cogredient with $\mathbf{A}$ (and satisfying the condition $\mathbf{A}\alpha\lambda = \alpha$). We use $N_{\alpha\lambda}$, N_α for the nerves of $\alpha\lambda$ and α, write $\sigma^{\beta\mu}_{\alpha\lambda}$ for the projections of $N_{\beta\mu}$ into $N_{\alpha\lambda}$, while σ^β_α denote now only those projections of N_β into N_α which are generated by a projection $\sigma^{\beta\mu}_{\alpha\lambda}$. As usual $K_{\alpha\lambda}$, K_α is the subcomplex of $N_{\alpha\lambda}$, N_α respectively formed by all simplices which have among their vertices at least one corresponding to an element of $\alpha\lambda$ (of α) with bicompact closure; $A_{\alpha\lambda}$ is the subcomplex of $K_{\alpha\lambda}$ defined by the vertices which correspond to elements of $\alpha\lambda$ meeting $\mathbf{A}$.

The passage from $\alpha\lambda$ to $\alpha = \mathbf{A}\alpha\lambda$ produces an isomorphic mapping(6) of

(6) Let N, N' be nerves, Q, Q' subcomplexes of N, N' respectively. We define $[Q] \subseteq N$ as the nerve consisting of all simplices of Q and of all faces of these simplices; the same for $[Q']$. Any (1-1)-mapping of Q on Q' generated by a (1-1)-simplicial mapping of $[Q]$ on $[Q']$ is called isomorphic.

$A_{\alpha\lambda}$ onto K_α and therefore an isomorphism $\phi_\alpha^{\alpha\lambda}$ of the groups $\Delta^r_{\alpha\lambda 0}$, $\nabla^r_{\alpha\lambda 0}$ (of $A_{\alpha\lambda}$) on the groups Δ^r_α, ∇^r_α (of K_α). The isomorphism $\phi_\alpha^{\alpha\lambda}$ satisfies the hypotheses of Lemma 33 and thus generates a well defined isomorphism

$$\text{of } \liminf (\Delta^r_{\alpha\lambda 0}, \varpi^{\beta\mu 0}_{\alpha\lambda 0}) \text{ on } \liminf (\Delta^r_\alpha, \varpi^\beta_\alpha);$$

$$\text{of } \limsup (\nabla^r_{\alpha\lambda 0}, \pi^{\alpha\lambda 0}_{\beta\mu 0}) \text{ on } \limsup (\nabla^r_\alpha, \pi^\alpha_\beta).$$

Since the coverings $\alpha\lambda$ form a cofinal part of the set of all coverings of $\mathbf{K}$ while the coverings α form a cofinal part of the set of all coverings of $\mathbf{A}$ we may identify

$$\liminf (\Delta^r_{\alpha\lambda 0}, \varpi^{\beta\mu 0}_{\alpha\lambda 0}) \text{ with } \Delta^r_A; \qquad \limsup (\nabla^r_{\alpha\lambda 0}, \pi^{\alpha\lambda 0}_{\beta\mu 0}) \text{ with } \nabla^r_A,$$

$$\liminf (\Delta^r_\alpha, \varpi^\beta_\alpha) \text{ with } \Delta^r_{\mathbf{A}}; \qquad \limsup (\nabla^r_\alpha, \pi^\alpha_\beta) \text{ with } \nabla^r_{\mathbf{A}}$$

and get in this way the *natural isomorphism* $(A\ \mathbf{A})$ of Δ^r_A, ∇^r_A on $\Delta^r_{\mathbf{A}}$, $\nabla^r_{\mathbf{A}}$; the inverse isomorphism is denoted by $(\mathbf{A}\ A)$; we return to them in section 37.

35. **The natural isomorphism between Δ^r_Γ, ∇^r_Γ and Δ^r_G, ∇^r_G.** Now we denote by α any covering of Γ, by $\alpha\lambda$ an arbitrary covering of $\mathbf{K}$ cogredient with Γ and such that $\Gamma\alpha\lambda = \alpha$. Then $N_{\alpha\lambda}$, N_α, $K_{\alpha\lambda}$, K_α have an obvious meaning. We retain among the inequalities $\beta\mu > \alpha\lambda$ only those for which there exists a projection $\sigma^{\beta\mu}_{\alpha\lambda}$ which assigns to each element $o_{\beta\mu} \in \beta\mu$ of the second kind an element $o_{\alpha\lambda} \in \alpha\lambda$, $o_{\alpha\lambda} \subseteq o_{\beta\mu}$, of the second kind. Only these $\sigma^{\beta\mu}_{\alpha\lambda}$ will be taken into account. In this way we get by virtue of 32.5 a cofinal part

$$K_1 = \{K_{\alpha\lambda}, \varpi^{\beta\mu}_{\alpha\lambda}\} \tag{35.1}$$

of the spectrum K.

According to section 32 we denote by $G'_{\alpha\lambda}$ the open subcomplex of $K_{\alpha\lambda}$, the simplices $T'_{\alpha\lambda} \in G'_{\alpha\lambda}$ being defined by the following property: among the vertices of $T_{\alpha\lambda}$ there is at least one corresponding to an inner element (of the second kind) of the covering $\alpha\lambda$. We denote by $\varpi'^{\beta\mu}_{\alpha\lambda}$ the projection of the cell complex $G'_{\beta\mu}$ into the cell complex $G'_{\alpha\lambda}$ generated by $\sigma^{\beta\mu}_{\alpha\lambda}$. It is easily seen [1, p. 88, formula (8.41)] that $K_\alpha = G'_{\alpha\lambda}$ for every $\alpha\lambda$; thus we may identify the group

$$\Delta^r_\alpha = \Delta^r_{K_\alpha} \text{ with } \Delta^r_{(\alpha\lambda)'} = \Delta^r_{G'_{\alpha\lambda}}; \qquad \nabla^r_\alpha = \nabla^r_{K_\alpha} \text{ with } \nabla^r_{(\alpha\lambda)'} = \nabla^r_{G'_{\alpha\lambda}}$$

and therefore

$$\Delta^r_\Gamma = \liminf (\Delta^r_\alpha, \varpi^\beta_\alpha) \text{ with } \Delta^r_{G'} = \liminf (\Delta^r_{(\alpha\lambda)'}, \varpi'^{\beta\mu}_{\alpha\lambda}),$$

$$\nabla^r_\Gamma = \limsup (\nabla^r_\alpha, \pi^\alpha_\beta) \text{ with } \nabla^r_{G'} = \limsup (\nabla^r_{(\alpha\lambda)'}, \pi'^{\alpha\lambda}_{\beta\mu}).$$

But since the coverings $\alpha\lambda$ (cogredient with Γ) on the one hand and the regular coverings on the other form both cofinal parts of the partially ordered set

of all coverings of $\mathbf{K}$ we get (taking into account 32.2) a natural isomorphism between

$$\Delta^r_{G'} \text{ and } \Delta^r_G; \qquad \nabla^r_{G'} \text{ and } \nabla^r_G,$$

and therefore between

$$\Delta^r_\Gamma \text{ and } \Delta^r_G; \qquad \nabla^r_\Gamma \text{ and } \nabla^r_G.$$

This isomorphism (considered from Γ to G) is denoted by $(\Gamma\ G)$, the inverse isomorphism by $(G\ \Gamma)$.

36. **The groups of the figure K, A, Γ and the results concerning them.** Because of the identities $\Delta^r_\mathbf{K}=\Delta^r_K$, $\nabla^r_\mathbf{K}=\nabla^r_K$ the natural isomorphisms transform the homomorphisms

$$E^A_K, \quad E^G_K, \quad J^K_A, \quad J^K_G$$

(of the groups Δ^r_A, ∇^r_G, ∇^r_K, Δ^r_K into the groups Δ^r_K, ∇^r_K, ∇^r_A, Δ^r_G) into the homomorphisms

$$E^\mathbf{A}_\mathbf{K}, \quad E^\Gamma_\mathbf{K}, \quad E^\mathbf{K}_\mathbf{A}, \quad J^\mathbf{K}_\Gamma$$

of the groups

$$\Delta^r_\mathbf{A}, \quad \nabla^r_\Gamma, \quad \nabla^r_\mathbf{K}, \quad \Delta^r_\mathbf{K}$$

into the groups

$$\Delta^r_\mathbf{K}, \quad \nabla^r_\mathbf{K}, \quad \nabla^r_\mathbf{A}, \quad \Delta^r_\Gamma$$

and thus define the groups

$$\Delta^r_{\mathbf{K}\mathbf{A}} = E^\mathbf{A}_\mathbf{K}\Delta^r_\mathbf{A} = E^A_K(\mathbf{A}A)\Delta^r_\mathbf{A} = E^A_K\Delta^r_A = \Delta^r_{KA}; \quad \nabla^r_{\mathbf{K}\Gamma} = E^\Gamma_\mathbf{K}\nabla^r_\Gamma = \nabla^r_{KG},$$

$$\Delta^r_{\mathbf{A}\mathbf{K}} = J^\mathbf{K}_\mathbf{A}\nabla^r_K = (A\mathbf{A})J^K_A\nabla^r_K = (A\mathbf{A})\nabla^r_{AK}; \quad \Delta^r_{\Gamma\mathbf{K}} = J^\mathbf{K}_\Gamma\Delta^r_\mathbf{K} = (G\Gamma)\Delta^r_{GK},$$

$$\Delta^r_{\mathbf{A}:\mathbf{K}} = (E^\mathbf{A}_\mathbf{K})^{-1}\,0_\mathbf{K} = (A\mathbf{A})(E^A_K)^{-1}\,0_K = (A\mathbf{A})\Delta^r_{A:K}; \quad \nabla^r_{\Gamma:\mathbf{K}} = (E^\Gamma_\mathbf{K})^{-1}\,0_\mathbf{K} = (G\Gamma)\nabla^r_{G:K},$$

$$\nabla^r_{\mathbf{A}:\mathbf{K}} = \nabla^r_\mathbf{A} - \nabla^r_{\mathbf{A}\mathbf{K}} = (A\mathbf{A})\nabla^r_A - (A\mathbf{A})\nabla^r_{AK} = (A\mathbf{A})\nabla^r_{A:K}; \quad \Delta^r_{\Gamma:\mathbf{K}}(G\Gamma)\Delta^r_{G:K}.$$

If we define the product of an element of $\Delta^r_\mathbf{A}$ (of ∇^r_Γ) with an element of $\nabla^r_\mathbf{A}$ (of Δ^r_Γ) as the product of the corresponding elements of Δ^r_A, ∇^r_A (of ∇^r_G, Δ^r_G), we see that not only all isomorphisms but also all dualities and annihilations between the groups of the figure K, A, G are transformed by the natural isomorphisms into the corresponding relations between the groups of the figure $\mathbf{K}$, $\mathbf{A}$, Γ. Therefore *all results* (14.1)–(14.3′) *formulated in section* 14 *for the figure* K, A, G *hold for* $\mathbf{K}$, $\mathbf{A}$, Γ.

37. **Remarks on the natural isomorphisms and the groups of the figure K, A, Γ.** Let an element of the group ∇^r_A (of the group Δ^r_A) be given. Take only the coverings $\alpha\lambda$ of $\mathbf{K}$ cogredient with $\mathbf{A}$. Using the subscript $\alpha\lambda 0$ for ∇-cycles (Δ-classes) on $A_{\alpha\lambda}$ we can write for any $z^r\in\nabla^r_A$, $\bar z^r\in\Delta^r_A$:

$$z^r = \{z^r_{\alpha\lambda 0}\}; \qquad \bar{z}^r = \{\mathfrak{Z}^r_{\alpha\lambda 0}\}.$$

As we know $A_{\alpha\lambda}$ can be identified with the nerve of $\alpha = \mathbf{A}\alpha\lambda$; therefore z^r, $\bar{z}^r$ can be considered as elements of $\nabla^r_{\mathbf{A}}$, $\Delta^r_{\mathbf{A}}$ and this identification represents an isomorphism—our natural isomorphism—$(A\ \mathbf{A})$.

For a moment, let us denote by α coverings of $\mathbf{\Gamma}$; let $z^r = \{z^r_\alpha\} \in \nabla^r_{\mathbf{\Gamma}}$. Each z^r_α can be considered as a ∇-cycle on $G'_{\alpha\lambda}$, thus ($G'_{\alpha\lambda} \subseteq G_{\alpha\lambda}$ being open on $G_{\alpha\lambda}$) as a ∇-cycle on $G_{\alpha\lambda}$, where $\alpha\lambda$ is cogredient with $\mathbf{\Gamma}$ and $\alpha = \mathbf{\Gamma}\alpha\lambda$. Conserve among the coverings of $\mathbf{K}$ only those which are cogredient with $\mathbf{\Gamma}$ and among the projections $\sigma^{\beta\mu}_{\alpha\lambda}$ of these coverings only projections which map elements of the second kind of $\beta\mu$ on elements of the second kind of $\alpha\lambda$. Then we can identify our z^r with a certain element of ∇^r_G and this identification is the natural isomorphism ($\mathbf{\Gamma}\ G$). We needed the regular coverings only to prove that this isomorphism maps $\nabla^r_{\mathbf{\Gamma}}$ onto ∇^r_G.

Now let α be a covering of $\mathbf{K}$. Let

$$\bar{z}^r_G = \{\mathfrak{Z}^r_{\alpha 1}\} \in \Delta^r_G.$$

Assign to any $\bar{z}^r_{\alpha 1} \in \mathfrak{Z}^r_{\alpha 1}$ the cycle

$$\bar{z}^r_{(\alpha)'} = J^{G'_\alpha}_{G^r} \bar{z}^r_{\alpha 1}$$

(this is a customary method if we consider $\bar{z}^r_{\alpha 1}$ as a Lefschetz relative cycle mod A_α and keep in mind that $A'_\alpha \supseteq A_\alpha$). We get in this way an isomorphism of Δ^r_G on $\Delta^r_{G'}$ (the "on" is again proved using the regular coverings). If now we take into account only the coverings of $\mathbf{K}$ cogredient with $\mathbf{\Gamma}$ we can identify $\bar{z}^r_{G'} = \{\bar{z}^r_{(\alpha)'}\} \in \Delta^r_{G'}$ with a certain element of $\Delta^r_{\mathbf{\Gamma}}$. Thus by the transition from Δ^r_G to $\Delta^r_{G'}$ and then to $\Delta^r_{\mathbf{\Gamma}}$ we get the natural isomorphism ($G\ \mathbf{\Gamma}$) of Δ^r_G onto $\Delta^r_{\mathbf{\Gamma}}$.

To the simple intuitive meaning of the natural isomorphisms corresponds a direct definition of the groups of the figure $\mathbf{K}$, $\mathbf{A}$, $\mathbf{\Gamma}$. Since $\Delta^r_{\mathbf{KA}} = \Delta^r_{KA}$, $\nabla^r_{\mathbf{K\Gamma}} = \nabla^r_{K\Gamma}$, $\nabla^r_{\mathbf{A}:\mathbf{K}} = \nabla^r_{\mathbf{A}} - \nabla^r_{\mathbf{AK}}$, $\Delta^r_{\mathbf{\Gamma}:\mathbf{K}} = \Delta^r_{\mathbf{\Gamma}} - \Delta^r_{\mathbf{\Gamma K}}$, we are interested in the groups $\nabla^r_{\mathbf{AK}}$, $\Delta^r_{\mathbf{\Gamma K}}$; $\Delta^r_{\mathbf{A}:\mathbf{K}}$, $\nabla^r_{\mathbf{\Gamma}:\mathbf{K}}$ only.

The elements of the bundle $z^r \in \nabla^r_A$ being identified with the ∇-cycles $z^r_{\alpha\lambda 0}$ on $A_{\alpha\lambda}$, $\alpha\lambda$ cogredient with $\mathbf{A}$, the bundles $z^r = \{z^r_{\alpha\lambda 0}\}$ belonging to $\nabla^r_{\mathbf{AK}}$ are characterized by the property: there is for each $z^r_{\alpha\lambda 0} \in z^r$ a $\beta\mu > \alpha\lambda$ such that $\pi^{\alpha\lambda 0}_{\beta\mu 0} z^r_{\alpha\lambda 0}$ is extensible over $K_{\beta\mu}$.

The elements $\bar{z}^r \in \Delta^r_{\mathbf{\Gamma K}}$ are characterized in the following way: we first identify the elements of the thread $\bar{z}^r$ with Δ-classes of $G'_{\alpha\lambda}$ ($\alpha\lambda$ cogredient with $\mathbf{\Gamma}$) and thus identify the thread $\bar{z}^r$ with a certain thread $\bar{z}'^r$ of the spectrum G'; then using regular coverings we identify the thread $\bar{z}'^r$ with a thread of the spectrum G; this latter must be extensible in the sense of section 26.

The elements of $\Delta^r_{\mathbf{A}:\mathbf{K}}$ are threads $\bar{z}^r \in \Delta^r_{\mathbf{A}}$ bounding on $\mathbf{K}$; it means that if we consider $\bar{z}^r_\alpha \in \mathfrak{Z}^r_\alpha \in \bar{z}^r \in \Delta^r_{\mathbf{A}:\mathbf{K}}$ as a Δ-cycle on $A_{\alpha\lambda}$ ($\alpha\lambda$ cogredient with $\mathbf{A}$, $\mathbf{A}\alpha\lambda = \alpha$), this cycle bounds on $K_{\alpha\lambda}$.

The elements of $\nabla^r_{\Gamma:\mathbf{K}}$ are bundles $z^r \in \nabla^r_\Gamma$ bounding on $\mathbf{K}$; it means that if we consider the elements of the bundle z^r as cycles on $G_{\alpha\lambda}$ ($\alpha\lambda$ cogredient with Γ), these cycles are elements of the bundle representing the zero of ∇^r_K. In other words, if $z^r_{\alpha\lambda} \in z^r$ and $z^r \in \nabla^r_\Gamma$, then there exists a $\beta\mu > \alpha\lambda$ such that $\pi^{\alpha\lambda}_{\beta\mu} z^r_{\alpha\lambda}$ bounds on $K_{\beta\mu}$.

As a matter of fact, the elements of the groups $\nabla^r_{\mathbf{A}:\mathbf{K}} = \nabla^r_{\mathbf{A}} - \nabla^r_{\mathbf{AK}}$ can be defined as *extension classes* of ∇-cycles: let α, β be two coverings of A; define N_α, N_β, K_α, K_β as in section 34; two ∇-cycles z^r_α, z^r_β on K_α, K_β respectively belong by definition to the same extension class if there exist coverings $\alpha\lambda$, $\beta\mu$, $\gamma\nu$ cogredient with $\mathbf{A}$ such that $\alpha = \mathbf{A}\alpha\lambda$, $\beta = \mathbf{A}\beta\mu$, $\gamma\nu > \alpha\lambda$, $\beta\mu$ and $\pi^{\alpha\lambda 0}_{\gamma\nu 0} z^r_\alpha - \pi^{\beta\mu 0}_{\gamma\nu 0} z^r_\beta$ be a ∇-cycle on $A_{\gamma\nu}$, extensible over $K_{\gamma\nu}$.

The elements of $\Delta^r_{\Gamma:\mathbf{K}}$ could be defined as *extension classes of threads*: two threads $\bar{z}^r = \{\mathfrak{Z}^r_\alpha\} \in \Delta^r_\Gamma$ and $\bar{z}'^r = \{\mathfrak{Z}''_\alpha\} \in \Delta^r_\Gamma$ belong by definition to the same extension class if for each α the cycle $z^r_\alpha - z''_\alpha$ (where $z^r_\alpha \in \mathfrak{Z}^r_\alpha$, $z''_\alpha \in \mathfrak{Z}''_\alpha$) considered as a cycle on $G'_{\alpha\lambda}$ ($\alpha\lambda$ cogredient with Γ and $\Gamma\alpha\lambda = \alpha$) is extensible over $K_{\alpha\lambda}$.

38. **Another form of the definition of the natural isomorphism** ($\mathbf{A}$ A). This other form will be used in Chapter IV, remark 1 in section 43. We conserve the notations of sections 30 and 31 and suppose for the sake of simplicity that $\mathbf{K}$ is bicompact. Take in each K_α the closed subcomplex A''_α with the same vertices as A_α and simplices $|e_{i_0} \cdots e_{i_r}|$ corresponding to those elements $o_{i_0}, \cdots, o_{i_r}$ of the covering α of $\mathbf{K}$ which satisfy the condition

$$\mathbf{A} \cap o_{i_0} \cap \cdots \cap o_{i_r} \neq 0.$$

Obviously A''_α is a subcomplex of A_α and

(38.1) $A''_\alpha = A_\alpha$ *for* α *cogredient with* $\mathbf{A}$.

Now denote by $\alpha\lambda$ the coverings of $\mathbf{K}$. To any $\alpha\lambda = \{o_1, \cdots, o_p, \cdots, o_s\}$ corresponds the covering

$$\alpha = \mathbf{A}\alpha\lambda = \{\mathbf{A}o_1, \cdots, \mathbf{A}o_p\}$$

of $\mathbf{A}$ where the $\mathbf{A}o_i = \mathbf{A} \cap o_i$ are "indexed sets" (in the sense of [1, p. 72, §5]), that is $\mathbf{A}o_i$ and $\mathbf{A}o_j$ are considered as different if $i \neq j$.

Each covering α of $\mathbf{A}$ can be put in the form $\alpha = \mathbf{A}\alpha\lambda$. The nerve of $\mathbf{A}\alpha\lambda$ being $A''_{\alpha\lambda}$ there is a natural isomorphism between $\nabla^r_{\mathbf{A}}$, $\Delta^r_{\mathbf{A}}$ and

$$\nabla^r_{A''} = \limsup (\nabla^r_{A''_{\alpha\lambda}}, \pi^{\alpha\lambda 0}_{\beta\mu 0}) \quad , \qquad \Delta^r_{A''} = \liminf (\Delta^r_{A''_{\alpha\lambda}}, \varpi^{\beta\mu 0}_{\alpha\lambda 0}).$$

By virtue of the identity (38.1) this isomorphism turns into an isomorphism between $\nabla^r_{\mathbf{A}}$, $\Delta^r_{\mathbf{A}}$ and ∇^r_A, Δ^r_A which is our natural isomorphism ($\mathbf{A}$ A).

§10. The case of continuous polyhedrons $\mathbf{K}$ and $\mathbf{A} \subseteq \mathbf{K}$

39. **An invariance theorem.** Let $\mathbf{K}$ be a finite polyhedron, K a simplicial decomposition (Simplizialzerlegung) of $\mathbf{K}$ (terminology of [4, pp. 128–129]).

Let A be a closed subcomplex of the complex K; then $\mathbf{A}$ denotes the polyhedron composed by the simplices of A; as usual $G=K\backslash A$, $\Gamma=\mathbf{K}\backslash\mathbf{A}$. Let us show that *the groups of the figure* $\mathbf{K}$, $\mathbf{A}$, Γ *are isomorphic to those of the figure* K, A, G. To this end denote by

$$K_1, K_2, \cdots, K_n, \cdots$$

the successive barycentric subdivisions of K; then the nth barycentric subdivision A_n of A is a subcomplex of K_n. Denote for a moment by λ'_n the covering of $\mathbf{K}$ by the *closed* barycentric stars of the vertices of K_n and take the open neighborhoods of the elements of λ'_n so closely as to get an open covering λ_n of $\mathbf{K}$ similar to λ'_n (that is, having the same nerve as λ'_n) and cogredient with $\mathbf{A}$. Put $\alpha_n=\mathbf{A}\lambda_n$. Obviously the λ_n and the α_n form cofinal parts of the sets of all coverings of $\mathbf{K}$ and of $\mathbf{A}$ respectively. The nerve of λ_n being K_n, and A_n being the nerve of α_n, the projections of K_{n+1} into K_n and of A_{n+1} into A_n map any vertex e_{n+1} of K_{n+1} (of A_{n+1}) on a vertex of the carrier of e_{n+1} in K_n (in A_n) and thus are "natural displacements" ("natürliche Verschiebungen" in the sense of [4, p. 349]). Therefore by a well known standard process the elements of $\Delta^r_{\mathbf{K}}$, $\Delta^r_{\mathbf{A}}$ are in (1-1)-correspondence with the elements of Δ^r_K, Δ^r_A, this correspondence realizing an isomorphism between

$$\Delta^r_{\mathbf{K}} \text{ and } \Delta^r_K, \qquad \Delta^r_{\mathbf{A}} \text{ and } \Delta^r_A,$$

$$\Delta^r_{\mathbf{KA}} \text{ and } \Delta^r_{KA}, \qquad \Delta^r_{\mathbf{A}:\mathbf{K}} \text{ and } \Delta^r_{A:K}.$$

All other groups of the figures $\mathbf{K}$, $\mathbf{A}$, Γ and K, A, G can be derived respectively from $\Delta^r_{\mathbf{K}}$, $\Delta^r_{\mathbf{KA}}$, $\Delta^r_{\mathbf{A}:\mathbf{K}}$, and Δ^r_K, Δ^r_{KA}, $\Delta^r_{A:K}$ by the same algebraic constructions; thus the groups of $\mathbf{K}$, $\mathbf{A}$, Γ are respectively isomorphic to the groups of K, A, G.

40. **The general polyhedral case.** Let $\mathbf{K}$ and $\mathbf{A}\subseteq\mathbf{K}$ be continuous polyhedrons (that is, topological images of finite polyhedrons). In this section coefficient-groups are the so called "elementary groups," that is discrete groups with a finite number of generators and their (bicompact) character groups. Any subgroup and any difference group of an elementary group being elementary, it follows that the groups $\Delta^r_{\mathbf{K}}$, $\nabla^r_{\mathbf{K}}$, $\Delta^r_{\mathbf{A}}$, $\nabla^r_{\mathbf{A}}$, $\Delta^r_{\mathbf{KA}}$, $\nabla^r_{\mathbf{K}\Gamma}$, $\Delta^r_{\mathbf{A}:\mathbf{K}}$, $\nabla^r_{\mathbf{AK}}$, $\nabla^r_{\mathbf{A}:\mathbf{K}}$, as well as (by the third pair of isomorphisms) the groups $\Delta^r_{\Gamma:\mathbf{K}}$, $\nabla^r_{\Gamma:\mathbf{K}}$ for $r\geqq 1$ are elementary. Now if—for a discrete group X—a subgroup U and the difference group $X-U$ are both elementary, then X itself is elementary (the proof is obvious, see for instance [4, p. 576, section 38]). By this lemma and the first pair of isomorphisms the groups Δ^r_Γ, ∇^r_Γ for $r\geqq 1$ are elementary. If X is the coefficient-group then, as is easily seen, $\nabla^0_\Gamma(X)$ is the direct sum of as many groups isomorphic to the group X as there are compact components of Γ. As in our case the number of these components is finite, ∇^0_Γ is elementary.

40.1. Thus if $\mathbf{K}$ and $\mathbf{A}$ are continuous polyhedrons, then with respect to

any elementary coefficient-group all groups of the figure $\mathbf{K}$, $\mathbf{A}$, $\mathbf{\Gamma}$ are elementary; they are all finite if the coefficient-group is finite.

For a finite coefficient-group X the order of the group $\nabla^r_{\mathbf{K}}$ is equal (by the second pair of isomorphisms) to the product of the orders of $\nabla^r_{\mathbf{K\Gamma}}$ and $\nabla^r_{\mathbf{AK}}$. This holds in particular for $X=I_m$ whatever the integer $m\geqq 2$. Since by a theorem of M. Bockstein [6] all groups $\nabla^r_{\mathbf{K}}(I)$ are completely determined by the orders of the groups $\nabla^r_{\mathbf{K}}(m)$ taken for all m and all r, it follows that all $\nabla^r_{\mathbf{K}}$ are determined by $\nabla^r_{\mathbf{K\Gamma}}(m)$ and $\nabla^r_{\mathbf{AK}}(m)$ (taken for all m and all r). In the same way by the first pair of isomorphisms the groups $\Delta^r_{\mathbf{A}}$ are determined by $\Delta^r_{\mathbf{A:K}}(m)$ and $\Delta^r_{\mathbf{KA}}(m)$ taken for all m and all r. To get an analogous result for the groups $\nabla^r_{\mathbf{\Gamma}}$, $\nabla^r_{\mathbf{\Gamma:K}}$, $\nabla^r_{\mathbf{K\Gamma}}$ we need the following

LEMMA. *The groups $\nabla^r_{\mathbf{\Gamma}}(m)$ (even the groups $\nabla^r_{\mathbf{\Gamma}}(X)$ for any X) can be deduced from the groups $\nabla^r_{\mathbf{\Gamma}}(I)$ (taken for all r) by means of the same relations as in the case where $\mathbf{\Gamma}$ is a finite cell complex.*

Let us assume for a moment that this lemma is proved. Then as the Bockstein theorem for a complex K rests solely upon the algebraic relations between $\nabla^r_K(I)$ and $\nabla^r_K(m)$, its original proof still holds for $\mathbf{\Gamma}$. Thus our lemma implies the following:

BOCKSTEIN THEOREM FOR $\mathbf{\Gamma}$. *The groups $\nabla^r_{\mathbf{\Gamma}}(I)$ are completely determined by the orders of the groups $\nabla^r_{\mathbf{\Gamma}}(m)$ taken for all m and all r.*

Now by the first pair of isomorphisms the order of $\nabla^r_{\mathbf{\Gamma}}(m)$ is equal to the product of the orders of $\nabla^r_{\mathbf{\Gamma:K}}(m)$ and $\nabla^r_{\mathbf{K\Gamma}}(m)$, and therefore $\nabla^r_{\mathbf{\Gamma}}(I)$ is determined by the groups $\nabla^r_{\mathbf{\Gamma:K}}(m)$ and $\nabla^r_{\mathbf{K\Gamma}}(m)$.

It remains only to prove the lemma. A proof of it is contained in the construction of Steenrod [13, section 11] which in our case gives the desired expression. But one can proceed more quickly by the following argument of Bockstein which uses only the final result of Steenrod and not his intermediate construction. The groups $\Delta^r_{\mathbf{\Gamma}}(I)$ being groups with a finite number of generators, they are respectively isomorphic to the groups $\nabla^r_N(I)$ where N is an appropriate nerve (see for example [4, p. 266, section 9]). Let $\mathbf{N}$ be the corresponding polyhedron. By the fundamental theorem of Steenrod the groups $\nabla^r_{\mathbf{N}}(m)$ can be expressed through the groups $\nabla^r_{\mathbf{N}}(I)$ and these expressions depend only on the groups involved (and the given coefficient-group, in our case I_m) and are therefore the same for $\mathbf{\Gamma}$ and $\mathbf{N}$. On the other hand, the groups $\nabla^r_{\mathbf{N}}(I)$, $\nabla^r_{\mathbf{N}}(m)$ are isomorphic to $\nabla^r_N(I)$, $\nabla^r_N(m)$ and the expression of $\nabla^r_{\mathbf{N}}(m)$ through the groups $\nabla^r_{\mathbf{N}}(I)$ is the same as the expression of $\nabla^r_N(m)$ through the groups $\nabla^r_N(I)$, which proves the lemma.

Thus all results announced in section 14 are now completely proved. Let us remark that the theorems of section 18 could be extended by the same methods first to spectra and then to spaces; such an extension does not present any serious difficulty and may be left to the reader.

CHAPTER IV. MANIFOLDS

§11. INFINITE COMPLEXES

41. **Infinite complexes**(7). Let Q be an infinite locally finite, complete (section 28), simplicial complex, thus a simplicial decomposition of a locally bicompact space (of an infinite polyhedron) $\mathbf{Q}$. Denote by L'_Q the group of the r-dimensional finite chains on Q with respect to the discrete coefficient-group X, by $\overline{L}'_Q$ the topologized (as in [13, p. 691]) group of all (infinite) r-dimensional chains on Q with respect to the bicompact coefficient-group $\Xi|X$. The topology in $\overline{L}'_Q$ is that which makes of $\overline{L}'_Q$ the direct sum of the groups $\overline{L}'_i$, the group $\overline{L}'_i \approx \Xi$ being the group of all chains of the form $\bar{a}t'_i$, where t'_i is an oriented simplex of Q and $\bar{a}$ runs over Ξ. Since L'_Q is the direct sum of the discrete groups $L'_i \approx X$ (the elements of L'_i being the chains at'_i with a running over X), the groups L'_Q and $\overline{L}'_Q$ are dual to each other. With these groups we get in the usual way the groups Δ'_Q and ∇'_Q: the group ∇'_Q (the group Δ'_Q) is the difference group of the group of all r-dimensional finite ∇-cycles (infinite Δ-cycles) over the subgroup of the bounding ones (a bounding ∇-cycle bounds by definition a finite chain).

The purpose of this section is the proof of the following theorem:

41.1. *If Q is a simplicial decomposition of the infinite polyhedron $\mathbf{Q}$, then*

$$\Delta'_{\mathbf{Q}} \approx \Delta'_Q; \qquad \nabla'_{\mathbf{Q}} \approx \nabla'_Q. \tag{41.1}$$

By virtue of the dualities $\Delta'_{\mathbf{Q}}|\nabla'_{\mathbf{Q}}$, $\Delta'_Q|\nabla'_Q$ it is sufficient to prove the first of the isomorphisms (41.1).

PRELIMINARY REMARKS. Let α be a (finite open) covering of $\mathbf{Q}$; an element of α (as well as the corresponding vertex of the nerve N_α) is called *regular* if its closure is bicompact; the other elements of α (and vertices of N_α) are called *special*. The simplices of N_α whose vertices are all special form the *special subcomplex* C_α of N_α; the open subcomplex $K_\alpha = N_\alpha \setminus C_\alpha$ is called the *regular part* of N_α; all regular vertices of N_α, and only the regular vertices, are zero-dimensional elements of K_α.

The open finite subcomplex G of Q is called a *regular subcomplex* of Q if it is the sum of the open stars(8) of some of the vertices of Q, or (which is the same) if each simplex of G has at least one vertex which belongs to G as an element. The proof of the following statement is obvious: If G is any open finite subcomplex of Q, while Q_1, G_1 are barycentric subdivisions of Q, G respectively, then G_1 is a regular subcomplex of Q_1.

Our last auxiliary definition in this section is the following: A covering α of Q is called *combinatorial with respect to* Q if the following conditions are satisfied:

(7) This section will not be needed in the rest of this paper.

(8) We use the word "star" for complexes as well as for the corresponding point sets.

(a) Each regular element of α is the open star of some vertex of Q.

(b) Let G_α be the (regular) subcomplex of Q defined as the sum of all open stars which are elements of α; let Γ_α be the corresponding point set in $\mathbf{Q}$ and $\overline{\Gamma}$ its closure (in $\mathbf{Q}$). Any special element of α either has no points in common with $\overline{\Gamma}$ or is of the form $o'_{\alpha i} \cup O_h$, where O_h is on $\mathbf{Q} \backslash \overline{\Gamma}$ and $o'_{\alpha i}$ is the open star (with respect to Q) of a vertex of some simplex of G_α.

LEMMA 41.2. *The regular part of the nerve N_α of any covering α combinatorial with respect to Q is a regular subcomplex of Q and any regular subcomplex G of Q is the regular part of the nerve of some covering combinatorial with respect to Q.*

The first statement of Lemma 41.2 follows immediately from the definition of a covering combinatorial with respect to Q. Let us prove the second statement. Let e_i be a vertex of an element of G. If $e_i \in G$ denote by o_i the open star of e_i; if e_i is not an element of G put $o_i = o'_i \cup (\mathbf{Q} \backslash \Gamma)$, where o'_i is the open star of e_i in Q. The open sets o_i form the required covering.

LEMMA 41.3. *Each covering α of Q has a refinement which is combinatorial with respect to a certain subdivision Q_λ of Q.*

Proof. Let $o_1, \cdots, o_p$ be the regular and $o_{p+1}, \cdots, o_s$ the special elements of α. Take the closed sets $a_i \subseteq o_i$ in such a way as to get a closed covering $\{a_1, \cdots, a_s\}$ of $\mathbf{Q}$. The a_i with $i \leqq p$ are bicompact while the $a_{p+1}, \cdots, a_s$ may be supposed to be not bicompact. Choose the subdivision Q_λ of Q such that any star of Q_λ meeting a_i is on o_i. Now let $o_{\lambda 1}, \cdots, o_{\lambda u}$ be all stars of vertices of Q_λ which meet any one of the sets $a_1, \cdots, a_p$. These stars form a regular subcomplex G_λ of Q_λ; denote by Γ_λ the corresponding subset of $\mathbf{Q}$. Denote by O_h, $h = p+1, \cdots, s$, the open sets composed by all stars of Q_λ which have points in common with a_h. Among all zero-dimensional elements of $Q_\lambda \backslash G_\lambda$ denote by $e'_1, \cdots, e'_q$ those which are vertices of elements of G_λ. Define for $i = 1, 2, \cdots, q$

$$o_{\lambda, u+i} = o'_{\lambda i} \cup (O_{h_i} \backslash \overline{\Gamma}_\lambda),$$

where $o'_{\lambda i}$ is the star of e'_i in Q_λ and $h_i \geqq p+1$ is chosen under the condition that a_{h_i} has points in common with $o'_{\lambda i}$. By the definition of O_h we have

$$o'_{\lambda i} \subseteq O_{h_i} \subseteq o_{h_i},$$

thus $o_{\lambda, u+i} \subseteq o_{h_i}$. If in the latter construction $O_{p+1}, \cdots, O_{p+q'}$ are the only sets O_h actually used, we put, for $j = q'+1, \cdots, s-p$,

$$o_{\lambda, u+q+j} = O_{p+j} \backslash \overline{\Gamma}_\lambda.$$

The open sets $o_{\lambda i}$, $i = 1, 2, \cdots, u+q+s-p$, form the required covering.

Now we go over to the proof of the isomorphism $\Delta'_{\mathbf{Q}} \approx \Delta'_{Q}$. Retain only those coverings of $\mathbf{Q}$ which are combinatorial with respect to subdivisions of Q. If α, β are combinatorial with respect to Q_λ, Q_μ then we set $\beta > \alpha$ if both of the following conditions are satisfied:

1. β is a refinement of α, α is not a refinement of β.
2. Q_μ is a subdivision of Q_λ.

The coverings with this ordering form a cofinal part of the set of all coverings of Q. The subcomplexes $K_\alpha = N_\alpha \backslash C_\alpha$ being open (even regular) subcomplexes of Q_λ, they undergo projections generated by natural displacements (section 39) of Q_μ into Q_λ. Let ϖ_α^β be those projections. Then

$$\Delta'_{\mathbf{Q}} = \liminf (\Delta'_{K_\alpha}, \varpi_\alpha^\beta). \tag{41.4}$$

Take on the other hand all finite open subcomplexes $G_{\alpha\lambda}$ of all subdivisions Q_λ of Q. We say that $G_{\beta\mu}$ follows on $G_{\alpha\lambda}$ and write $\beta\mu > \alpha\lambda$ if Q_μ is a subdivision of Q_λ and the subdivision $s_\mu^\lambda G_{\alpha\lambda}$ of $G_{\alpha\lambda}$ in Q_μ is a subcomplex of $G_{\beta\mu}$. In particular, if G_α are the subcomplexes of Q itself, then $\beta > \alpha$ means that $G_\alpha \subset G_\beta$. Denote by σ_λ^μ the natural displacements of Q_μ into Q_λ and write (as indices) $(\beta\mu)\lambda$ instead of $\sigma_\lambda^\mu G_{\beta\mu}$ and $\alpha\lambda$ instead of $G_{\alpha\lambda}$. Since $G_{\alpha\lambda}$ is an open subcomplex of $\sigma_\lambda^\mu G_{\beta\mu}$ we may define a projection $\rho_{\alpha\lambda}^{\beta\mu}$ of the cell complex $G_{\beta\mu}$ into $G_{\alpha\lambda}$ by means of

$$\rho_{\alpha\lambda}^{\beta\mu} x_{\beta\mu} = J_{\alpha\lambda}^{(\beta\mu)\lambda} \sigma_\lambda^\mu x_{\beta\mu}$$

for any chain $x_{\beta\mu}$ on $G_{\beta\mu}$. This gives for $\mu = \lambda$ the relation $\rho_{\alpha\lambda}^{\beta\lambda} x_{\beta\lambda} = J_{\alpha\lambda}^{\beta\lambda} x_{\beta\lambda}$ and in particular $\rho_\alpha^\beta x_\beta = J_\alpha^\beta x_\beta$ on Q.

Those $G_{\alpha\lambda}$ which are subdivisions of appropriate $G_\alpha \subset Q$ form a cofinal part of the set of all $G_{\alpha\lambda}$, and $\rho_\alpha^{\alpha\lambda}$ maps isomorphically $\Delta'_{\alpha\lambda}$ on Δ'_α. Therefore $\liminf (\Delta'_{\alpha\lambda}, \rho_{\alpha\lambda}^{\beta\mu}) \approx \liminf (\Delta'_\alpha, J_\alpha^\beta)$. But the groups $\liminf (\Delta'_\alpha, J_\alpha^\beta)$ and $\Delta'_{\mathbf{Q}}$ are isomorphic as shown by Steenrod [13, pp. 691–692]; thus

$$\Delta'_{\mathbf{Q}} \approx \liminf (\Delta'_{\alpha\lambda}, \rho_{\alpha\lambda}^{\beta\mu}). \tag{41.5}$$

On the other hand, since the regular subcomplexes form a cofinal part in the set of all $G_{\alpha\lambda}$, we deduce from Lemma 41.2 the isomorphism

$$\Delta'_{\mathbf{Q}} \approx \liminf (\Delta'_{\alpha\lambda}, \rho_{\alpha\lambda}^{\beta\mu}). \tag{41.6}$$

From (41.5) and (41.6) follows $\Delta'_{Q} \approx \Delta'_{\mathbf{Q}}$, q.e.d.

42. **The groups δ'_Q and the duality theorem of Alexander-Pontrjagin.** Let X be a discrete coefficient-group. Denote by δ'_Q the difference group of the group of all r-dimensional finite Δ-cycles of the (infinite) complex Q over the subgroup of the bounding cycles (a bounding cycle bounds by definition a finite chain). In this section we suppose that Q is an n-dimensional combinatorial (h-)manifold (finite or infinite), that is, that for the open star O_i of any vertex e_i of Q

$$\Delta^n_{O_i}(I) \approx \nabla^n_{O_i}(I) \approx I,$$

while for $r<n$ the groups $\Delta^r_{O_i}(I) \approx \nabla^r_{O_i}(I)$ contain the zero element only. We suppose moreover that Q is orientable, which means that for every component Q_i of Q, $\Delta^n_{Q_i}(I) \approx \nabla^n_{Q_i}(I) \approx I$. The oriented (open) barycentric stars of Q form a cell complex Q^*. If t^p_i is an oriented simplex of Q, then we denote by τ^q_i, with $q=n-p$, the barycentric star with the center in the center of t^p_i oriented in such a way that the intersection number $(t^p_i \times \tau^q_i)$ be equal to $+1$ (see for example [12, §69]). Then we have in the cell complex Q^* the incidence numbers

$$(\tau^{q+1}_j : \tau^q_i) = (-1)^p (t^p_i : t^{p-1}_j).$$

Define now for any p-chain x^p on Q the q-chain $D^q x^p$ on Q^* taking on τ^q_i the same value as x^p takes on t^p_i. An easy calculation shows that

$$\Delta D^q x^p = (-1)^{p+1} D^{q-1} \nabla x^p \tag{42.1}$$

whence it follows that the operator D^q generates an isomorphism between ∇^p_Q and $\delta^q_{Q^*}$. As both δ^q_Q and $\delta^q_{Q^*}$ are isomorphic to $\delta^q_{Q_1}$, where Q_1 is the barycentric subdivision of Q, we have the

DUALITY THEOREM OF POINCARÉ.

$$\nabla^p_Q \approx \delta^q_Q, \qquad q = n - p.$$

Let $\mathbf{Q}$ be an n-dimensional continuous manifold, that is, a finite or infinite polyhedron, one and thus all simplicial decompositions of which are combinatorial manifolds. Then any open set on $\mathbf{Q}$ is also a continuous manifold (see [4, pp. 143–146, Theorem "of Runge"]). There follows then from the duality theorems of Kolmogoroff and of Poincaré the

DUALITY THEOREM OF ALEXANDER-PONTRJAGIN. *If the n-dimensional orientable continuous manifold $\mathbf{Q}$ is simply connected in the dimensions p and $p+1$, $\mathbf{A}$ is a closed set on $\mathbf{Q}$, and $\Gamma=\mathbf{Q}\setminus\mathbf{A}$, then the groups $\Delta^p_{\mathbf{A}}$ and δ^{q-1}_Γ are dual (the groups $\nabla^p_{\mathbf{A}}$ and δ^{q-1}_Γ are isomorphic) to each other. Here the coefficient-group for δ^{q-1}_Γ and $\nabla^p_{\mathbf{A}}$ is the discrete group X while the coefficient-group for $\Delta^p_{\mathbf{A}}$ is the bicompact group $\Xi \mid X$.*

The generalization of this theorem for an arbitrary orientable manifold (not necessarily simply connected in any dimension) is given by the

GENERAL ALEXANDER-PONTRJAGIN DUALITY THEOREM. *For any n-dimensional orientable continuous manifold $\mathbf{K}$, any closed set $\mathbf{A}\subseteq\mathbf{K}$, and $\Gamma=\mathbf{K}\setminus\mathbf{A}$, the group $\nabla^p_{\mathbf{A}:\mathbf{K}}$ is isomorphic to the group $\delta^{q-1}_{\Gamma:\mathbf{K}}$, where the latter group is defined as the difference group of the r-dimensional finite cycles on Γ over the subgroup of the cycles bounding on Γ (coefficient-group discrete).*

Here the finite cycles on Γ are to be understood in some invariant sense (as "continuous cycles" in the sense of [4, pp. 333–339] or as convergent cycles on compact sets $\phi \subset \Gamma$ (section 43)) or to be replaced by the δ-classes of section 43.

To prove this theorem we have to show only that the groups $\nabla^r_{\Gamma:\mathbf{K}}$ and $\delta^r_{\Gamma:\mathbf{K}}$ are isomorphic. This can be done by the standard methods of proving invariance theorems (see for example [4, Chapter IX] and [3]); these methods, however, so far as the writer sees, need at a certain moment the Combinatorial Lemma 45.3. But the same lemma is the only essential point in proving directly the isomorphism $\nabla^p_{\mathbf{A}:\mathbf{K}} \approx \delta^{q-1}_{\Gamma:\mathbf{K}}$. We prefer therefore to give this direct proof which rests entirely on section 15 (Chapter I), being practically independent of Chapters II and III.

§12. The groups $\Delta^r_{\mathbf{A}:\mathbf{K}}$, $\nabla^r_{\mathbf{A}:\mathbf{K}}$, and $\delta^r_{\Gamma:\mathbf{K}}$ for a manifold $\mathbf{K}$

43. The groups $\Delta^r_{\mathbf{A}:\mathbf{K}}$ and $\nabla^r_{\mathbf{A}:\mathbf{K}}$ for a polyhedron $\mathbf{K}$. Let $\mathbf{K}$ be a polyhedron; we suppose this polyhedron finite (that is bicompact) although the following definitions hold for infinite polyhedrons too. As usual $\mathbf{A}$ is any closed set in $\mathbf{K}$ and $\Gamma = K \setminus A$. In this case the groups $\Delta^r_{\mathbf{A}:\mathbf{K}}$, $\nabla^r_{\mathbf{A}:\mathbf{K}}$ can be defined in a quite elementary way, independent of the definitions and results of Chapters II and III. Let

$$K_1, K_2, \cdots, K_\alpha, \cdots \tag{43.1}$$

be a sequence of successive subdivisions of a given simplicial subdivision K of $\mathbf{K}$; we suppose moreover that for any integer $\alpha \geqq 1$ the complex $K_{\alpha+1}$ is the result of one or more successive barycentric subdivisions of K_α. Taking in K_α all simplices which meet $\mathbf{A}$ and all faces of these simplices we get a closed subcomplex A_α of K_α. Put in correspondence to any vertex $e_{\alpha+1}$ of $K_{\alpha+1}$ a fixed vertex $e_\alpha = \sigma^{\alpha+1}_\alpha e_{\alpha+1}$ of the carrier of $e_{\alpha+1}$ in K_α. This gives a well defined simplicial mapping (a "natural displacement," section 39) called a *projection* of $K_{\alpha+1}$ onto K_α. The projection $\sigma^{\alpha+1}_\alpha$ maps $A_{\alpha+1}$ into A_α. For $\beta > \alpha+1$ put

$$\overset{\beta}{\sigma_\alpha} = \overset{\alpha+1}{\sigma_\alpha}\ \overset{\alpha+2}{\sigma_{\alpha+1}} \cdots \overset{\beta}{\sigma_{\beta-1}}$$

which implies the transitivity relation $\sigma^\gamma_\alpha = \sigma^\beta_\alpha \sigma^\gamma_\beta$ for any $\gamma > \beta > \alpha$.

The projection σ^β_α generates a homomorphism ϖ^β_α of $L^r_\beta = L^r_{K_\beta}$ into $L^r_\alpha = L^r_{K_\alpha}$ and a homomorphism $\varpi^{\beta 0}_{\alpha 0}$ of $L^r_{\beta 0} = L^r_A$ into $L^r_{\alpha 0} = L^r_{A_\alpha}$; the conjugate homomorphisms π^α_β and $\pi^{\alpha 0}_{\beta 0}$ of the group L^r_α into L^r_β and of $L^r_{\alpha 0}$ into $L^r_{\beta 0}$ are defined by

$$(\pi^\alpha_\beta x^r_\alpha \cdot t^r_\beta) = (x^r_\alpha \cdot \overset{\beta}{\sigma_\alpha} t^r_\beta)$$

for any chain x^r_α on K_α (on A_α) and any oriented simplex t^r_β of K_β (of A_β). The homomorphisms $\varpi^{\beta 0}_{\alpha 0}$, $\pi^{\alpha 0}_{\beta 0}$ generate homomorphisms (denoted also by $\varpi^{\beta 0}_{\alpha 0}$, $\pi^{\alpha 0}_{\beta 0}$) of $\Delta^r_{\beta 0}$ into $\Delta^r_{\alpha 0}$ (of $\nabla^r_{\alpha 0}$ into $\nabla^r_{\beta 0}$) and map $\Delta^r_{\beta 0;\beta} = \Delta^r_{A_\beta:K_\beta}$ into $\Delta^r_{\alpha 0:\alpha} = \Delta^r_{A_\alpha:K_\alpha}$; $\nabla^r_{\alpha 0:\alpha} = \nabla^r_{A_\alpha:K_\alpha}$ into $\nabla^r_{\beta 0;\beta} = \nabla^r_{A_\beta:K_\beta}$. We define now:

$$\Delta^r_{\mathbf{A}:\mathbf{K}} = \lim\inf\,(\Delta^r_{\alpha 0:\alpha},\ \varpi^{\beta 0}_{\alpha 0}); \qquad \nabla^r_{\mathbf{A}:\mathbf{K}} = \lim\sup\,(\nabla^r_{\alpha 0:\alpha},\ \pi^{\alpha 0}_{\beta 0}).$$

REMARK 1. It is easily seen that the groups $\Delta^r_{\mathbf{A}:\mathbf{K}}$, $\nabla^r_{\mathbf{A}:\mathbf{K}}$ just defined do not depend upon the special choice of the sequence (43.1). But one can also show that the definitions of this section agree with those of Chapter III. In fact, the simplices of A_α are characterized among the simplices of K_α by the property that their open stars meet $\mathbf{A}$; therefore A_α is the nerve of the covering

$$\alpha = \{o_{\alpha 1}, \cdots, o_{\alpha s}\}$$

of $\mathbf{A}$, where $o_{\alpha i} = \mathbf{A} \cap O_{\alpha i}$ and the $O_{\alpha i}$ are those open stars of (the vertices of) K_α which meet $\mathbf{A}$. In other words, if $\alpha\lambda$ is the covering of $\mathbf{K}$ by the open stars of the vertices of K_α then in the sense of section 38 we have $\alpha = \mathbf{A}\alpha\lambda$. The coverings $\alpha\lambda$, $\alpha = 1, 2, \cdots$, form a confinal part of the set of all coverings of $\mathbf{K}$ and the projections σ^β_α put in correspondence to each $O_{\beta j}$ an $O_{\alpha i} \supseteq O_{\beta j}$. If we take into account section 38 we see at once that the definition (43.2) gives us the same groups as the definitions of Chapter III.

REMARK 2. The elements of $\Delta^r_{\mathbf{A}:\mathbf{K}}$ are those "threads" or sequences

$$z^r = (\overset{r}{\gamma}_1, \overset{r}{\gamma}_2, \cdots, \overset{r}{\gamma}_\alpha, \cdots) \in \Delta^r_{\mathbf{A}}, \qquad \overset{r}{\gamma}_\alpha \in \Delta^r_{\alpha 0}, \qquad \varpi^{\beta 0}_{\alpha 0}\overset{r}{\gamma}_\beta = \overset{r}{\gamma}_\alpha$$

whose elements $\overset{r}{\gamma}_\alpha \in \Delta^r_{\alpha 0}$ are classes of cycles on A_α bounding on K_α.

It is easy to see that the group $\Delta^r_{\mathbf{A}:\mathbf{K}}$ can also be defined as the difference group of the group of all r-dimensional convergent cycles (see for instance [5, p. 241]) of the set $\mathbf{A}$ over the subgroup of cycles bounding on $\mathbf{K}$.

The elements of $\nabla^r_{\mathbf{A}:\mathbf{K}}$ are *extension classes* of r-dimensional ∇-cycles; the ∇-cycle z^r_α on A_α and the ∇-cycle z^r_β on A_β belong by definition to the same extension class if there exists a $\gamma > \alpha, \beta$ such that $\pi^\alpha_\gamma z^r_\alpha - \pi^\beta_\gamma z^r_\beta$ is extensible over K_γ.

REMARK 3. The coefficient-groups in $\Delta^r_{\mathbf{A}:\mathbf{K}}$, $\nabla^r_{\mathbf{A}:\mathbf{K}}$ are dual: bicompact for $\Delta^r_{\mathbf{A}:\mathbf{K}}$ and discrete for $\nabla^r_{\mathbf{A}:\mathbf{K}}$.

44. **The groups $\delta^r_{\Gamma:\mathbf{K}}$ for a manifold $\mathbf{K}$.** From now to the end of this paper $\mathbf{K}$ is a continuous n-dimensional manifold (definition in section 42) which for sake of simplicity we suppose closed (that is bicompact) although all following definitions and results hold without this restriction. We conserve the notations of section 43 and denote by $K_{\alpha 1}$ the barycentric subdivision of K_α (thus $K_{\alpha+1}$ either coincides with $K_{\alpha 1}$ or is a subdivision of $K_{\alpha 1}$). The complex of all (open) barycentric stars of K_α is denoted by K^*_α. The barycentric stars which correspond to simplices of $G_\alpha = K_\alpha \backslash A_\alpha$ form a closed subcomplex G^*_α of K^*_α. Since the elements of K^*_α are subcomplexes of $K_{\alpha 1}$, we get a *closed subcomplex* $G^*_{\alpha 1}$ of $K_{\alpha 1}$, called the *barycentric subdivision of* G^*_α, if we take all simplices of $K_{\alpha 1}$ lying on elements T^*_α of G^*_α (that is all simplices which are elements of the subcomplexes $T^*_\alpha \in G^*_\alpha$). The polyhedron composed of all simplices of $G^*_{\alpha 1}$ is denoted by $\Gamma^*_{\alpha 1}$. It is $\Gamma^*_{\alpha 1} \subseteq \Gamma$ and even $\Gamma = \cup^\infty_{\alpha=1} \Gamma^*_{\alpha 1}$. Without loss of generality we may suppose that $\Gamma_{\alpha 1} \subseteq \Gamma_{(\alpha+1)1}$. Therefore the subdivision

of $G^*_{\alpha 1}$ into simplices of $K_{\beta 1}$, $\beta \geqq \alpha$, is a subcomplex $s^{\alpha 1}_{\beta 1} G^*_{\alpha 1}$ of $G^*_{\beta 1}$, and for any chain $x^r_{\alpha 1}$ on $G_{\alpha 1}$ we have the subdivision $s^{\alpha 1}_{\beta 1} x^r_{\alpha 1}$ on $G^*_{\beta 1}$.

REMARK 1. By definition

$$(s^{\alpha 1}_{\beta 1} x^r_{\alpha 1} \cdot t^r_{\beta 1}) = (x^r_{\alpha 1} \cdot t^r_{\alpha 1})$$

if the simplex $t^r_{\beta 1}$ of $G^*_{\beta 1}$ lies on $t^r_{\alpha 1}$ and its orientation is coherent with the orientation of $t^r_{\alpha 1}$; $(s^{\alpha 1}_{\beta 1} x^r_{\alpha 1} \cdot t^r_{\beta 1}) = 0$ if $t^r_{\beta 1}$ lies on a higher dimensional element of $G^*_{\alpha 1}$.

REMARK 2. In an analogous way we define the subdivision $s^{\alpha}_{\beta 1} x^r_{\alpha}$ of a chain x^r_{α} on G^*_{α}. Obviously for $\beta > \alpha$

$$s^{\alpha}_{\beta 1} x^r_{\alpha} = s^{\alpha 1}_{\beta 1} s^{\alpha}_{\alpha 1} x^r_{\alpha}.$$

After these preliminary remarks we define the group $\delta^r_{\Gamma:K}$ as follows. We say that two Δ-cycles, the cycle $z^r_{\alpha 1}$ on $G^*_{\alpha 1}$ and the cycle $z^r_{\beta 1}$ on $G^*_{\beta 1}$, belong to the same *r-dimensional δ-class on* Γ if there is a $\gamma > \alpha$, β such that $s^{\alpha 1}_{\gamma 1} z^r_{\alpha 1} - s^{\beta 1}_{\gamma 1} z^r_{\beta 1}$ bounds on $G^*_{\gamma 1}$. To get the *sum of two δ-classes* z^r and z'^r take ad libitum $z^r_{\alpha 1} \in z^r$, $z^r_{\beta 1} \in z'^r$ and define $z^r + z'^r$ as the δ-class containing $s^{\alpha 1}_{\gamma 1} z^r_{\alpha 1} + s^{\beta 1}_{\gamma 1} z^r_{\beta 1}$ where γ is any integer which is greater than or equal to the greatest among the integers α and β.

DEFINITION 44.1. *The discrete additive group of all r-dimensional δ-classes on Γ is called the group δ^r_Γ (with respect to the given discrete coefficient-group).*

DEFINITION 44.2. *The subgroup $\delta^r_{\Gamma:K}$ of δ^r_Γ is formed by the δ-classes $z^r \in \delta^r_\Gamma$ whose elements $z^r_{\alpha 1}$ bound on (the corresponding) $K_{\alpha 1}$.*

REMARK 3. The groups δ^r_Γ, $\delta^r_{\Gamma:K}$ could be defined also as follows. A convergent cycle on a compact subset of Γ (with respect to a discrete coefficient-group X) is called a cycle on Γ; by definition, it bounds on Γ if it bounds on a compact subset of Γ. The difference group of the group of all r-dimensional cycles on Γ over the subgroup of the cycles bounding on Γ is the group δ^r_Γ; the subgroup of δ^r_Γ whose elements are classes of cycles bounding on K is the group $\delta^r_{\Gamma:K}$; the latter group can also be defined as the difference group of the group of all r-dimensional cycles on Γ over the subgroup of the cycles bounding on K.

§13. THE GENERAL ALEXANDER-PONTRJAGIN DUALITY THEOREM: REDUCTION TO THE COMBINATORIAL LEMMA

45. **Formulation of the theorem and of the lemma.** From now on we either suppose the manifold K orientable or take for the coefficient-group the group I_2. All notations of §12 hold.

The theorem to be proved is:

45.1. *The groups $\Delta^p_{A:K}$ and $\delta^{n-1}_{\Gamma:K}$ $(q = n-p)$ are dual;*

or (equivalent formulation):

45.2. *The groups* $\nabla^p_{A:K}$ *and* $\delta^{q-1}_{\Gamma:K}$ $(q=n-p)$ *are isomorphic.*

The proof rests upon the following

COMBINATORIAL LEMMA 45.3. *If* z^p_α *is a* ∇*-cycle on* K_α *equal to zero on* A_α *then for any* $\beta>\alpha$ *we have*

$$s^{\beta}_{\beta 1}D^q_\beta \pi^{\alpha}_{\beta} z^p_\alpha \sim s^{\alpha}_{\alpha 1}D^q_\alpha z^p_\alpha \text{ on } G^*_{\beta 1} \tag{45.3}$$

($\sim$ *is the sign of homology, section* 9, D^q *is the duality operator of section* 42).

This lemma will be proved in the next paragraph; in the present paragraph we suppose the lemma proved and prove on this basis the Theorem 45.2. We write E_α instead of $E^{A_\alpha}_{K_\alpha}$.

46. **The operators $\Delta D^q_\alpha E_\alpha$ and the isomorphism $\Delta D^q E$.** To each ∇-cycle $z^p_{\alpha 0}$ on A_α corresponds the ∇-cycle $\nabla E_\alpha z^p_{\alpha 0}$ on G_α and thus the Δ-cycle $D^{q-1}_\alpha \nabla E_\alpha z^p_{\alpha 0} = (-1)^{p+1}\Delta D^q_\alpha E_\alpha z^p_{\alpha 0}$ on G^*_α bounding on K^*_α. If a ∇-cycle z^p_α on G_α bounds on G_α the chain x^{p-1}_α, then $D^q_\alpha z^p_\alpha$ bounds on G^*_α the chain $\pm D^{q+1}_\alpha x^{p-1}_\alpha$ and vice versa. Therefore from section 15 we have:

46.1. *The operator* $\Delta D^q_\alpha E_\alpha$ *produces an isomorphism of the group* $\nabla^p_{A_\alpha:K_\alpha}$ *onto the group* $\Delta^{q-1}_{G^*_\alpha:K_\alpha *}$.

Our purpose is to show that the operators $\Delta D^q_\alpha E_\alpha$ taken for different α produce an isomorphism—we call it $\Delta D^q E$—of $\nabla^p_{A:K}$ onto $\delta^{q-1}_{\Gamma:K}$. We need for this proof some minor lemmas.

LEMMA 46.2. *For any* ∇*-cycle* $z^p_{\alpha 0}$ *on* A_α *and* $\beta>\alpha$

$$\pi^{\alpha}_{\beta}\nabla E_\alpha z^p_{\alpha 0} \sim \nabla E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0} \text{ on } G_\beta. \tag{46.2}$$

Proof. As π^α_β commutes with ∇ we have to prove the homology

$$\nabla \pi^{\alpha}_{\beta} E_\alpha z^p_{\alpha 0} \sim \nabla E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0} \text{ on } G_\beta$$

or

$$\nabla(\pi^{\alpha}_{\beta} E_\alpha z^p_{\alpha 0} - E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0}) \sim 0 \text{ on } G_\beta.$$

But an easy direct calculation shows that the chain $\pi^\alpha_\beta E_\alpha z^p_{\alpha 0} - E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0}$ is itself on G_β, which proves the assertion.

The fundamental lemma follows:

LEMMA 46.3. *For any* ∇*-cycle* $z^p_{\alpha 0}$ *on* A_α *and* $\beta>\alpha$

$$s^{\beta}_{\beta 1}\Delta D^q_\beta E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0} \sim s^{\alpha}_{\alpha 1}\Delta D^q_\alpha E_\alpha z^p_{\alpha 0} \text{ on } G^*_{\beta 1}. \tag{46.3}$$

Proof. Set in (45.3)

$$z^{p+1}_\alpha = \nabla E_\alpha z^p_{\alpha 0}$$

instead of z^p_α; it becomes

$$s^{\beta}_{\beta 1} D^{q-1}_{\beta} \pi^{\alpha}_{\beta} \nabla E_\alpha z^p_{\alpha 0} \sim s^{\alpha}_{\beta 1} D^{q-1}_{\alpha} \nabla E_\alpha z^p_{\alpha 0} \text{ on } G^*_{\beta 1}. \tag{46.31}$$

Apply to both sides of (46.2) the operator D^{q-1}_β; this gives

$$D^{q-1}_{\beta} \pi^{\alpha}_{\beta} \nabla E_\alpha z^p_{\alpha 0} \sim D^{q-1}_{\beta} \nabla E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0} \text{ on } G^*_\beta.$$

Thus on the left-hand side of (46.31) we can replace $D^{q-1}_\beta \pi^\alpha_\beta \nabla E_\alpha z^p_{\alpha 0}$ by $D^{q-1}_\beta \nabla E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0}$ and get

$$s^{\beta}_{\beta 1} D^{q-1}_{\beta} \nabla E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0} \sim s^{\alpha}_{\beta 1} D^{q-1}_{\alpha} \nabla E_\alpha z^p_{\alpha 0} \text{ on } G^*_{\beta 1}. \tag{46.32}$$

But by (42.1)

$$\begin{cases} D^{q-1}_{\beta} \nabla E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0} = (-1)^{p+1} \Delta D^q_\beta E_\beta \pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0}, \\ D^{q-1}_{\alpha} \nabla E_\alpha z^p_{\alpha 0} = (-1)^{p+1} \Delta D^q_\alpha E_\alpha z^p_{\alpha 0}. \end{cases}$$

If we bring this into (46.32) we get (46.3).

It is now easy to prove that the homomorphisms $\Delta D^q_\alpha E_\alpha$ taken for different α generate a homomorphism $\Delta D^q E$ of $\nabla^p_{\mathbf{A}:\mathbf{K}}$ into $\delta^{q-1}_{\Gamma:\mathbf{K}}$: since every $\Delta D^q_\alpha E_\alpha z^p_{\alpha 0}$ bounds on K^*_α we need only to verify the following

LEMMA 46.4. *If for the ∇-cycles $z^p_{\alpha 0}$ on A_α and $z^p_{\beta 0}$ on A_β there exists a $\gamma > \alpha, \beta$ such that $\pi^{\alpha 0}_{\gamma 0} z^p_{\alpha 0} - \pi^{\beta 0}_{\gamma 0} z^p_{\beta 0}$ is extensible over K_γ, then*

$$s^{\alpha}_{\gamma 1} \Delta D^q_\alpha E_\alpha z^p_{\alpha 0} \sim s^{\beta}_{\gamma 1} \Delta D^q_\beta E_\beta z^p_{\beta 0} \text{ on } G^*_{\gamma 1}. \tag{46.4}$$

Proof. From 46.1 it follows that $\Delta D^q_\gamma E_\gamma (\pi^{\alpha 0}_{\gamma 0} z^p_{\alpha 0} - \pi^{\beta 0}_{\gamma 0} z^p_{\beta 0})$ bounds on G^*_γ, thus

$$s^{\gamma}_{\gamma 1} \Delta D^q_\gamma E_\gamma \pi^{\alpha 0}_{\gamma 0} z^p_{\alpha 0} \sim s^{\gamma}_{\gamma 1} \Delta D^q_\gamma E_\gamma \pi^{\beta 0}_{\gamma 0} z^p_{\beta 0} \text{ on } G^*_{\gamma 1}. \tag{46.41}$$

On the other hand (46.3) gives (with β replaced by γ)

$$s^{\gamma}_{\gamma 1} \Delta D^q_\gamma E_\gamma \pi^{\alpha 0}_{\gamma 0} z^p_{\alpha 0} \sim s^{\alpha}_{\gamma 1} \Delta D^q_\alpha E_\alpha z^p_{\alpha 0} \text{ on } G^*_{\gamma 1}$$

and (with α, β replaced by β, γ respectively)

$$s^{\gamma}_{\gamma 1} \Delta D^q_\gamma E_\gamma \pi^{\beta 0}_{\gamma 0} z^p_{\beta 0} \sim s^{\beta}_{\gamma 1} \Delta D^q_\beta E_\beta z^p_{\beta 0} \text{ on } G^*_{\gamma 1}.$$

These homologies substituted into (46.41) give (46.4).

46.5. The homomorphism $\Delta D^q E$ of $\nabla^p_{\mathbf{A}:\mathbf{K}}$ into $\delta^{q-1}_{\Gamma:\mathbf{K}}$ is an isomorphism. This follows from

LEMMA 46.51. *If we have for a ∇-cycle $z^p_{\alpha 0}$ on A_α and a certain $\beta > \alpha$ the homology*

$$s^{\alpha}_{\beta 1} \Delta D^q_\alpha E_\alpha z^p_{\alpha 0} \sim 0 \text{ on } G^*_{\beta 1},$$

then $\pi^{\alpha 0}_{\beta 0} z^p_{\alpha 0}$ is extensible over K_β.

Proof. By (46.3)

$$s^{\alpha}_{\beta 1}\Delta D^{q}_{\alpha}E_{\alpha}z^{p}_{\alpha 0} \sim s^{\beta}_{\beta 1}\Delta D^{q}_{\beta}E_{\beta}\pi^{\alpha 0}_{\beta 0}z^{p}_{\alpha 0} \text{ on } G^{*}_{\beta 1},$$

thus $s^{\beta}_{\beta 1}\Delta D^{q}_{\beta}E_{\beta}\pi^{\alpha 0}_{\beta 0}z^{p}_{\alpha 0}$ bounds on $G^{*}_{\beta 1}$. It is known (see for instance [4, p. 247, Theorem III]): If a cycle of the form $s^{\beta}_{\beta 1}z^{*}_{\beta}$, where z^{*}_{β} is a Δ-cycle on G^{*}_{β}, bounds on $G^{*}_{\beta 1}$, then z^{*}_{β} bounds on G^{*}_{β}. Therefore $\Delta D^{q}_{\beta}E_{\beta}\pi^{\alpha 0}_{\beta 0}z^{p}_{\alpha 0}$ bounds on G^{*}_{β} and by 46.1 the ∇-cycle $\pi^{\alpha 0}_{\beta 0}z^{p}_{\alpha 0}$ is extensible.

46.6. The isomorphism $\Delta D^{q}E$ maps $\nabla^{p}_{\mathbf{A}:\mathbf{K}}$ onto $\delta^{q-1}_{\Gamma:\mathbf{K}}$. Let z^{q-1} be an element of $\delta^{q-1}_{\Gamma:\mathbf{K}}$. It is known (see for instance [4, p. 246, Theorem III]): Any Δ-cycle on $G^{*}_{\alpha 1}$ is homologous on $G^{*}_{\alpha 1}$ to a cycle of the form $s^{\alpha}_{\alpha 1}z^{*}_{\alpha}$ with z^{*}_{α} on G^{*}_{α}. It follows that the δ-class z^{q-1} contains a cycle $s^{\alpha}_{\alpha 1}z^{*}_{\alpha}$, where z^{*}_{α} is a Δ-cycle on G^{*}_{α} and bounds on K^{*}_{α}. From 46.1 it follows that there exists a ∇-cycle $z^{p}_{\alpha 0}$ on A_{α} such that $\Delta D^{q}_{\alpha}E_{\alpha}z^{p}_{\alpha 0}$ is homologous to z^{*}_{α} on G^{*}_{α} and therefore

$$s^{\alpha}_{\alpha 1}\Delta D^{q}_{\alpha}E_{\alpha}z^{p}_{\alpha} \in z^{q-1},$$

which proves our assertion.

§14. Proof of the combinatorial lemma

47. **The fundamental identity.** We write in this paragraph T for non-oriented and t for oriented simplices. The simplex $T^{r}_{\alpha 1}$ of $K_{\alpha 1}$ whose vertices are the centers of the simplices $T^{n_0}_{\alpha} > \cdots > T^{n_r}_{\alpha}$ will be denoted by

$$T^{r}_{\alpha 1} = \left| T^{n_0}_{\alpha} > \cdots > T^{n_r}_{\alpha} \right|.$$

A simplex $T^{r}_{\alpha 1}$ of $K_{\alpha 1}$ is called *principal* if its dimension is the same as the dimension of its carrier in K^{*}_{α}; the other simplices of $K_{\alpha 1}$ are called *accessory*. Among all simplices of $K_{\alpha 1}$ the principal simplices and only these are of the form

$$T^{q}_{\alpha 1} = \left| T^{n}_{\alpha} > T^{n-1}_{\alpha} > \cdots > T^{p}_{\alpha} \right|.$$

If we put in correspondence to the center of each simplex T_{β} of K_{β} the center of the simplex $\sigma^{\beta}_{\alpha}T_{\beta}$ we get a simplicial mapping $\sigma^{\beta 1}_{\alpha 1}$ of $K_{\beta 1}$ into $K_{\alpha 1}$: for $T^{r}_{\beta 1} = \left| T^{n_0}_{\beta} > T^{n_1}_{\beta} > \cdots > T^{n_r}_{\beta} \right| \in K_{\beta 1}$ we get

$$\sigma^{\beta 1}_{\alpha 1}T^{r}_{\beta 1} = \left| \sigma^{\beta}_{\alpha}T^{n_0}_{\beta} \geqq \sigma^{\beta}_{\alpha}T^{n_1}_{\beta} \geqq \cdots \geqq \sigma^{\beta}_{\alpha}T^{n_r}_{\beta} \right| \in K_{\alpha 1}.$$

Remark 1. If σ^{β}_{α} does not degenerate on T^{n}_{β} (that is, if the dimensions of $\sigma^{\beta}_{\alpha}T^{n}_{\beta}$ and T^{n}_{β} are the same) then for any simplex $T^{q}_{\beta 1}$ lying on T^{n}_{β} the simplex $\sigma^{\beta 1}_{\alpha 1}T^{q}_{\beta 1}$ is the image of $T^{q}_{\beta 1}$ under the affine transformation σ^{β}_{α} of T^{n}_{β}.

The manifold $\mathbf{K}$ being orientable, choose any definite orientation of K_{α} and the corresponding orientation of K_{β} as well as of all n-dimensional simplices T^{n}_{α} and T^{n}_{β} of K_{α} and K_{β}. These orientations will be denoted by t^{n}_{α}, t^{n}_{β}. For $p<n$ take ad libitum a definite orientation of each p-dimensional simplex $T^{p}_{\alpha} \in K_{\alpha}$ and denote this orientation by t^{p}_{α}. Choose the orientation τ^{q}_{α} of the

corresponding barycentric star $\mathbf{T}^q_\alpha$ in such a way as to get the intersection number $(t^p_\alpha \times \tau^q_\alpha) = +1$. For $T^q_{\alpha 1} \in \mathbf{T}^q_\alpha$ take the orientation $t^q_{\alpha 1}$ corresponding to τ^q_α.

Choose as follows the orientations t^p_β of the simplices T^p_β of K_β. If σ^β_α does not degenerate on T^p_β, and $\sigma^\beta_\alpha T^p_\beta = T^p_\alpha$, define t^p_β by the condition $\sigma^\beta_\alpha t^p_\beta = t^p_\alpha$. If σ^β_α degenerates on T^p_β, choose t^p_β arbitrarily. The orientation τ^q_β of the barycentric star $\mathbf{T}^q_\beta$ corresponding to T^p_β is again defined by the condition $(t^p_\beta \times \tau^q_\beta) = +1$. Finally, the orientation $t^q_{\beta 1}$ of $T^q_{\beta 1} \in \mathbf{T}^q_\beta$ corresponds to τ^q_β.

The proof of the combinatorial lemma 45.3 rests on the following

FUNDAMENTAL IDENTITY.

$$\sigma^{\beta 1}_{\alpha 1} s^{\beta}_{\beta 1} D^q_\beta \pi^\alpha_\beta x^p_\alpha = s^{\alpha}_{\alpha 1} D^q_\alpha x^p_\alpha \tag{47.1}$$

for any chain x^p_α on K_α and any $\beta > \alpha$.

REMARK 2. The mapping σ^β_α (of K_β on K_α), being a natural displacement, has the degree 1 (see [4, p. 348, Theorem II]). If σ^β_α is an arbitrary simplicial mapping of the degree d of the arbitrary n-dimensional closed orientable manifold on K_β on the arbitrary n-dimensional closed manifold K_α, then (47.1) is replaced by

$$\sigma^{\beta 1}_{\alpha 1} s^{\beta}_{\beta 1} D^q_\beta \pi^\alpha_\beta x^p_\alpha = d s^{\alpha}_{\alpha 1} D^q_\alpha x^p_\alpha \tag{47.2}$$

for any chain x^p_α on K_α where the operator π^α_β is defined as usual by $(\pi^\alpha_\beta x^p_\alpha \cdot t^p_\beta) = (x^p_\alpha \cdot \sigma^\beta_\alpha t^p_\beta)$ for any oriented $t^p_\beta \in K_\beta$. In this general form the identity (47.2) formulates a known property of the Hopf "Umkehrhomomorphismus" (see for instance [7]). The proof given below involves this general case.

48. **Proof of the fundamental identity.** Let us see how the principal simplices are transformed under $\sigma^{\beta 1}_{\alpha 1}$. First of all:

48.1. *If the accessory q-dimensional simplex*

$$T^q_{\alpha 1} = |\, T^{n_0}_\alpha > \cdots > T^{n_q}_\alpha | \in K_{\alpha 1}$$

is the image under $\sigma^{\beta 1}_{\alpha 1}$ of the principal simplex $T^q_{\beta 1} = |\, T^n_\beta > T^{n-1}_\beta > \cdots > T^p_\beta |$, then $n_q < p$.

Proof. From $\sigma^\beta_\alpha T^{n-i}_\beta = T^{n_i}_\alpha$ follows $n_i \leqq n-i$, in particular $n_0 \leqq n$, $n_q \leqq p$. If $n_q = p$, the simplex $T^q_{\alpha 1}$ would be principal, and 48.1 is proved.

The *principal simplices* $T^q_{\beta 1}$ mapped under $\sigma^{\beta 1}_{\alpha 1}$ on the *principal* $T^q_{\alpha 1} = |\, T^n_\alpha > \cdots > T^p_\alpha |$ are easily seen to be in (1-1)-correspondence with the $T^n_\beta \in K_\beta$ mapped under σ^β_α on T^n_α. Moreover the orientations t^n_α, t^n_β, t^p_α, t^p_β, $t^q_{\alpha 1}$, $t^q_{\beta 1}$ have been chosen in such a way that from $T^q_{\beta 1} = |\, T^n_\alpha > \cdots > T^p_\alpha |$ and

$$\sigma^{\beta}_{\alpha} t^n_\beta = \epsilon t^n_\alpha \qquad (\text{with } \epsilon = \pm 1) \tag{48.11}$$

follows

$$\sigma_{\alpha 1}^{\beta 1}\, {}^{q}t_{\beta 1}^{q} = \epsilon\, {}^{q}t_{\alpha 1}^{q} \tag{48.12}$$

with the same ϵ in (48.11) and (48.12) (the simple proof uses remark 1 of section 47). Therefore for any principal simplex $t_{\alpha 1}^{q}$

$$(\sigma_{\alpha 1}^{\beta 1} \sum t_{\beta 1}^{q} \cdot t_{\alpha 1}^{q}) = (\sigma_{\alpha}^{\beta} \sum t_{\beta}^{n} \cdot t_{\alpha}^{n}),$$

where the sum is taken over all principal $t_{\beta 1}^{q}$ at the left and over all n-dimensional simplices of K_β at the right. Since the value of $\sigma_\alpha^\beta \sum t_\beta^n$ on t_α^n is equal to the degree d we get

$$\sigma_{\alpha 1}^{\beta 1} \sum t_{\beta 1}^{q} = d \sum t_{\alpha 1}^{q} \tag{48.2}$$

(summation on both sides over all q-dimensional principal simplices of $K_{\beta 1}$ and $K_{\alpha 1}$ respectively).

The proof of the fundamental identity presents no more difficulty. To simplify the notations put

$$y_{\beta 1}^{q} = {}^{\beta} s_{\beta 1} D_{\beta}^{q}\, {}^{\alpha}\pi_{\beta}\, {}^{p}x_{\alpha}.$$

We shall calculate the value of the chain $y_{\beta 1}^{q}$ on an arbitrary simplex $t_{\beta 1}^{q}$ of $K_{\beta 1}$. If $t_{\beta 1}^{q}$ is accessory this value is zero by the very definition of the operator $s_{\beta 1}^{\beta}$. If $t_{\beta 1}^{q}$ is a principal simplex, let τ_β^q be its carrier in K_β^* and t_β^p the corresponding (oriented) simplex of K_β. By the definition of $s_{\beta 1}^{\beta}$, D_β^q and π_β^α we have

$$(y_{\beta 1}^{q} \cdot t_{\beta 1}^{q}) = (x_\alpha^p \cdot \sigma_\alpha^\beta t_\beta^p). \tag{48.3}$$

To evaluate the left-hand side of the identity (47.2) take any $T_{\alpha 1}^{q} \in K_{\alpha 1}$ and all principal simplices

$$T_j^q = \left| T_\beta^n(j) > \cdots > T_\beta^p(j) \right| \in K_{\beta 1}, \qquad j = 1, 2, \cdots, \nu,$$

which are mapped by $\sigma_{\alpha 1}^{\beta 1}$ on $T_{\alpha 1}^{q}$. Let

$$\sigma_{\alpha 1}^{\beta 1}\, {}^{q}t_j^q = \epsilon_j\, {}^{q}t_{\alpha 1}^{q}, \qquad \epsilon_j = \pm 1.$$

We have by (48.3) and the definition of the ϵ_j:

$$(\sigma_{\alpha 1}^{\beta 1}\, {}^{q}y_{\beta 1}^{q} \cdot t_{\alpha 1}^{q}) = \sum_j (y_{\beta 1}^{q} \cdot \epsilon_j t_j^q) = \sum_j \epsilon_j (x_\alpha^p \cdot \sigma_\alpha^\beta t_\beta^p(j)). \tag{48.4}$$

If $T_{\alpha 1}^{q}$ is accessory then by (48.1) we have $n_q < p$; it follows—since $\sigma_\alpha^\beta T_\beta^n(j) = T_\alpha^{n_q}$—that $(x_\alpha^p \cdot \sigma_\alpha^\beta t_\beta^p(j)) = 0$, therefore by (48.4)

$$(\sigma_{\alpha 1}^{\beta 1}\, {}^{q}y_{\beta 1}^{q} \cdot t_{\alpha 1}^{q}) = 0.$$

Obviously $(s_{\alpha 1}^{\alpha} D_\alpha^q x_\alpha^p \cdot t_{\alpha 1}^{q}) = 0$, thus both sides of (47.2) take on the accessory $t_{\alpha 1}^{q}$ the value zero.

If $T_{\alpha 1}^{q}$ is principal, $T_{\alpha 1}^{q} = \left| T_\alpha^n > \cdots > T_\alpha^p \right|$, then $\sigma_\alpha^\beta T_\beta^p(j) = T_\alpha^p$, and even $\sigma_\alpha^\beta t_\beta^p(j) = t_\alpha^p$, hence by (48.4)

$$(\sigma_{\alpha 1}^{\beta 1} y_{\beta 1}^{q} \cdot t_{\alpha 1}^{q}) = \sum_j \epsilon_j (x_\alpha^p \cdot t_\alpha^p) = (x^p \cdot t_\alpha^p) \sum_j \epsilon_j.$$

By (48.2) $\sum_j \epsilon_j = d$ and therefore

$$(\sigma_{\alpha 1}^{\beta 1} y_{\beta 1}^{q} \cdot t_{\alpha 1}^{q}) = d(x_\alpha^p \cdot t_\alpha^p).$$

So much for the left-hand side of (47.2). At the right, the chain $s_{\alpha 1}^{\alpha} D_\alpha^q x_\alpha^p$ takes the value $(x_\alpha^p \cdot t_\alpha^p)$ on $t_{\alpha 1}^q$, which completes the proof of the identity (47.2).

49. **The rest of the proof of Lemma** (45.3). This is based on the following two elementary propositions:

49.1. Let Γ_α, $\Gamma_{\alpha 1}^*$ denote the set theoretical sum of the simplices of G_α, $G_{\alpha 1}^*$ respectively. There exists a continuous mapping C of the point set Γ_α onto the point set $\Gamma_{\alpha 1}^*$ leaving all points of $\Gamma_{\alpha 1}^*$ invariant.

49.2. Let the ∇-cycle z_α^p on K_α be equal to zero on A_α; if $z_{\beta 1}^q = s_{\beta 1}^{\beta} D_\beta^q \pi_\beta^\alpha z_\alpha^p$ differs from zero on a given $T_{\beta 1}^q \in K_{\beta 1}$ then a convex subset of Γ_α contains both the closures of $T_{\beta 1}^q$ and of $\sigma_{\alpha 1}^{\beta 1} T_{\beta 1}^q$.

Proof of 49.1. The open set Γ_α is the set theoretical sum of the simplices $T_{\alpha 1} \in K_{\alpha 1}$ lying on it. Each of the simplices $T_{\alpha 1} \in K_{\alpha 1}$ lying on Γ_α but not belonging to $G_{\alpha 1}^*$ is of the form

$$T_{\alpha 1}^r = |T_\alpha^{n_0} > \cdots > T_\alpha^{n_i} > \cdots > T_\alpha^{n_r}|,$$

where $T_\alpha^{n_0}, \cdots, T_\alpha^{n_i}$ belong to G_α, while $T_\alpha^{n_{i+1}}, \cdots, T_\alpha^{n_r}$ are elements of A_α. Thus each point a of $T_{\alpha 1}^r$ is on a segment $|a'a''| \subseteq T_{\alpha 1}^r$ joining a point a' of the simplex $T_{\alpha 1}' = |T_\alpha^{n_0} > \cdots > T_\alpha^{n_i}| \in G_{\alpha 1}^*$ with a point a'' of $T_{\alpha 1} = |T_\alpha^{n_{i+1}} > \cdots > T_\alpha^{n_r}|$, the simplex $T_{\alpha 1}''$ belonging to the barycentric subdivision $A_{\alpha 1}$ of A_α. If we put $C(a) = a'$ for $a \in \Gamma_\alpha \setminus \Gamma_{\alpha 1}^*$ and $C(a) = a$ for $a \in \Gamma_{\alpha 1}^*$ we obtain a continuous mapping C with the required properties.

Proof of 49.2. Let T^n be a simplex, T^p a face of T^n. Denote by $S(T^p, T^n)$ the complex composed of T^n and all faces T^r of T^n satisfying the condition $T^r \geqq T^p$ (where $T' > T''$ means that the simplex T'' is a face of T'). Denote by $\sum(T^p, T^n)$ the set theoretical sum of all elements of $S(T^p, T^n)$. Among all points of the closed set $\overline{T}^n$ the points of $\sum(T^p, T^n)$ are characterized by the property that their barycentric coordinates (in the coordinate system defined by the vertices of T^n) corresponding to the vertices of T^p are all positive. Therefore the point set $\sum(T^p, T^n)$ is convex. This obvious remark will be applied now to the proof of 49.2. The hypotheses of 49.2 imply that $T_{\beta 1}^q$ is a principal simplex,

$$T_{\beta 1}^q = |T_\beta^n > \cdots > T_\beta^p|,$$

and that $\pi_\beta^\alpha z_\alpha^p$ differs from zero on T_β^p; hence

$$\sigma_\alpha^\beta T_\beta^p = T_\alpha^p \in G_\alpha.$$

Denote by T_α^n the carrier of T_β^n in K_α, and put

$$T_\alpha^{n_0} = \sigma_\alpha^\beta T_\beta^n.$$

The mapping σ_α^β being a natural displacement, we have

$$T_\alpha^{n_0} \leqq T_\alpha^n$$

and therefore

$$T_\alpha^p = \sigma_\alpha^\beta T_\beta^p \leqq T_\alpha^n.$$

Since $T_\alpha^p \in G_\alpha$ and G_α is open in K_α, we have

$$S(T_\alpha^p, T_\alpha^{n_0}) \subseteq S(T_\alpha^p, T_\alpha^n) \subseteq G_\alpha.$$

The vertices of $\sigma_{\alpha 1}^{\beta 1} T_{\beta 1}^q$, being the centers of the simplices

$$T_\alpha^{n_0} = \sigma_\alpha^\beta T_\beta^n \geqq \sigma_\alpha^\beta T_\beta^{n-1} \geqq \cdots \geqq \sigma_\alpha^\beta T_\beta^p = T_\alpha^p,$$

are points of the set $\sum(T_\alpha^p, T_\alpha^{n_0}) \subseteq \sum(T_\alpha^p, T_\alpha^n)$. Since the point set $\sum(T_\alpha^p, T_\alpha^n)$ is convex and contains the vertices of the simplex $\sigma_{\alpha 1}^{\beta 1} T_{\beta 1}^q$ it contains the closure of this simplex. It remains to prove that the closure of the simplex $T_{\beta 1}^q$ is contained in the same convex set $\sum(T_\alpha^p, T_\alpha^n)$. Let $T_\alpha^{n_i}$, $i = 1, 2, \cdots, q$, be the carrier of T_β^{n-i} in K_α. Since T_α^n is the carrier of T_β^n and $T_\alpha^p = \sigma_\alpha^\beta T_\beta^p \leqq T_\alpha^{n_q}$, we have

$$T_\alpha^n \geqq T_\alpha^{n_1} \geqq \cdots \geqq T_\alpha^{n_q} \geqq T_\alpha^p,$$

whence it follows that $T_\alpha^n, T_\alpha^{n_1}, \cdots, T_\alpha^{n_q}$ and therefore the centers of $T_\beta^n \subseteq T_\alpha^n$, $T_\beta^{n-1} \subseteq T_\alpha^{n_1}, \cdots, T_\beta^p \subseteq T_\alpha^{n_q}$ are on $\sum(T_\alpha^p, T_\alpha^n)$; since $\sum(T_\alpha^p, T_\alpha^n)$ is convex, the closure of the simplex $T_{\beta 1}^q = |T_\beta^n > T_\beta^{n-1} > \cdots > T_\beta^p|$ is a subset of $\sum(T_\alpha^p, T_\alpha^n)$ and 49.2 is proved.

It follows from 49.2 that the prism with the bases $z_{\beta 1}^q$ and $\sigma_{\alpha 1}^{\beta 1} z_{\beta 1}^q$ (see section 28) lies on a polyhedron $\Pi \subseteq \Gamma_\alpha$ (composed of the simplices of this prism) and thus $z_{\beta 1}^q \sim \sigma_{\alpha 1}^{\beta 1} z_{\beta 1}^q$ on $\Pi \subseteq \Gamma_\alpha$ (this homology being understood in some invariant sense—either as a continuous homology in the sense of [4, p. 335] or as a homology between convergent cycles [5, pp. 240, 241]). If we take $z_{\beta 1}^q$ from (49.2) we obtain

$$(49.3) \qquad {}^{\beta}s_{\beta 1} D_\beta^q \pi_\beta^\alpha z_\alpha^p \sim \sigma_{\alpha 1}^{\beta 1} {}^{\beta}s_{\beta 1} D_\beta^q \pi_\beta^\alpha z_\alpha^p \text{ on } \Pi \subseteq \Gamma_\alpha \subseteq \Gamma_\beta.$$

On the other hand, we have by (47.1)

$${}^{\alpha}s_{\alpha 1} D_\alpha^q z_\alpha^p = \sigma_{\alpha 1}^{\beta 1} {}^{\beta}s_{\beta 1} D_\beta^q \pi_\beta^\alpha z_\alpha^p;$$

thus ${}^{\alpha}s_{\alpha 1} D_\alpha^q z_\alpha^p$ (considered as a continuous or as a convergent cycle) lies on Π and therefore

$${}^{\alpha}s_{\beta 1} D_\alpha^q z_\alpha^p = {}^{\alpha 1}s_{\beta 1} {}^{\alpha}s_{\alpha 1} D_\alpha^q z_\alpha^p \sim {}^{\alpha}s_{\alpha 1} D_\alpha^q z_\alpha^p \text{ on } \Pi.$$

Hence we may replace in (49.3) the chain $\sigma_{\alpha 1}^{\beta 1} {}^{\beta}s_{\beta 1} D_\beta^q \pi_\beta^\alpha z_\alpha^p$ by ${}^{\alpha}s_{\beta 1} D_\alpha^q z_\alpha^p$, which gives

$$s^{\beta}_{\beta 1} D^{q}_{\beta} \pi_{\beta}{}^{\alpha} z^{p}_{\alpha} \sim s^{\alpha}_{\beta 1} D^{q}_{\alpha} z^{p}_{\alpha} \text{ on } \Pi \subseteq \Gamma_{\beta}.$$

By the mapping C of 49.1 this homology turns into

$$s^{\beta}_{\beta 1} D^{q}_{\beta} \pi_{\beta}{}^{\alpha} z^{p}_{\alpha} \sim s^{\alpha}_{\beta 1} D^{q}_{\alpha} z^{p}_{\alpha} \text{ on } \Gamma_{\beta 1}, \text{ thus on } G^{*}_{\beta 1},$$

which completes the proof of the combinatorial lemma, as well as the proof of the general duality theorem of Alexander-Pontrjagin.

§15. The removing theorem

50. **The removing theorem in the case of a manifold.** A ∇-cycle z^{p}_{α} on K_{α} is homologous on K_{α} to a cycle on G_{α} if and only if the Δ-cycle $D^{q}_{\alpha} z^{p}_{\alpha}$ is homologous on K^{*}_{α} to a cycle on G^{*}_{α}. As the scalar product of z^{p}_{α} with a Δ-cycle $z^{p}_{\alpha 0}$ on A_{α} is equal to the intersection number of $D^{q}_{\alpha} z^{p}_{\alpha}$ with the same $z^{p}_{\alpha 0}$ we get the following form of

THE REMOVING THEOREM FOR MANIFOLDS. *A δ-class z^{q} on the (n-dimensional orientable) manifold $\mathbf{K}$ contains as a subset a δ-class on the given open set $\mathbf{\Gamma} = \mathbf{K} \setminus \mathbf{A}$ if and only if, whatever be the thread $z^{p}_{0} = \{\iota^{p}_{\delta 0}\} \in \Delta^{p}_{\alpha}$ and $z^{q}_{\alpha} \in z^{q}$, the intersection of z^{q}_{α} with any $z^{p}_{\alpha 0} \in \iota^{p}_{\delta 0}$ is equal to zero (here, as usual, the δ-classes are taken with respect to a discrete coefficient-group X, while the coefficient-group in Δ^{p}_{α} is $\Xi | X$).*

Another formulation of the same theorem is:

A q-dimensional convergent cycle z^{q} (with respect to the discrete group X) on $\mathbf{K}$ is homologous on $\mathbf{K}$ to a cycle on $\mathbf{\Gamma}$ if and only if the intersection of z^{q} with each p-dimensional convergent cycle on $\mathbf{A}$ (with respect to $\Xi | X$) is equal to zero.

REMARK 1. The theorem holds also for $X = \Xi = \mathfrak{R}$ (see section 5).

REMARK 2. For non-orientable manifolds the theorem holds in the special case $X = \Xi = I_{2}$.

Bibliography

1. P. Alexandroff, *General combinatorial topology*, Trans. Amer. Math. Soc. vol. 49 (1941) pp. 41–105.

2. ———, (a) *Homologiegruppen allgemeiner Projectionsspectra*; (b) *Allgemeiner Dualitätssatz für Projectionsspectra und für im Kleinen bikompakte Räume*, Mitteilungen der Georgischen Abteilung der Akademie der Wissenschaften der SSR vol. 2 (1941) pp. 213–219 and 315–319 (in Russian with a German summary).

3. ———, *Zurückfürung des Alexander-Pontrjaginschen Dualitätssatzes auf den Dualitätssatz von Kolmogoroff*, ibid. vol. 1 (1940) pp. 401–410.

4. P. Alexandroff and H. Hopf, *Topologie*, I, 1935.

5. P. Alexandroff, H. Hopf, L. Pontrjagin, *Über den Brouwerschen Dimensionsbegriff*, Compositio Math. vol. 4 (1935) pp. 239–255.

6. M. Bockstein, *Über die Homologiegruppen der Vereinigung zweier Komplexe*, Rec. Math. (Mat. Sbornik) N.S. vol. 9 (1941) pp. 365–376.

7. H. Freudenthal, *Zum Hopfschen Umkehrhomomorphismus*, Ann. of Math. (2) vol. 38 (1937) pp. 847–853.

8. A. Kolmogoroff, *Über die Dualität im Aufbau der kombinatorischen Topologie*, Rec. Math. (Mat. Sbornik) N.S. vol. 1 (1936) pp. 97–102.

9. S. Lefschetz, *On closed sets on a manifold*, Ann. of Math. (2) vol. 29 (1928) pp. 232–254.

10. L. Pontrjagin, *Topological groups*.

11. ———, *Zum Alexanderschen Dualitätssatz*. II, Nachr. Ges. Wiss. Göttingen, 1927, pp. 446–454.

12. Seifert-Threlfall, *Topologie*, 1934.

13. N. E. Steenrod, *On universal homology groups*, Amer. J. Math. vol. 48 (1936) pp. 661–701.

14. A. W. Tucker, *An abstract approach to manifolds*, Ann. of Math. (2) vol. 34 (1935) pp. 191–243.

MATHEMATICAL INSTITUTE OF THE ACADEMY OF SCIENCES,
KASAN, USSR, January–June, 1942

INDEPENDENT INTEGRAL BASES AND A CHARACTERIZATION OF REGULAR SURFACES

BY

H. T. MUHLY[1]

Introduction. The fundamental paper *Theorie der algebraischen Funktionen einer Veränderlichen* of Dedekind and Weber opened a wide field of research. The methods which these authors brought into play in the study of algebraic functions of one variable have lent themselves readily, from the conceptual point of view at least, to generalization to the case of several variables. However, even if one restricts himself to the case of algebraic functions of two variables he finds a sharp line of demarcation in the analogy with the one variable case when he attacks problems of enumeration such as the problem of Riemann-Roch.

Several years ago Zariski made the conjecture that the analogy between the theory of algebraic curves and the theory of algebraic surfaces would extend much further in the case of regular surfaces, that is surfaces whose arithmetic and geometric genera coincide. The objective of this paper is to establish the truth of this conjecture and to show that this fact is in a sense characteristic of regular surfaces.

A tool which proved to be most useful to Dedekind and Weber in deriving the Riemann-Roch theorem as well as other allied results was furnished by their so-called *normal bases*, the construction of which we shall describe briefly below.

Let Σ be a field of algebraic functions of one variable x over a field of constants K, which is assumed to be maximally algebraic in Σ and of characteristic zero. The field Σ is then an algebraic extension of $K(x)$ of finite relative degree ν. Let $\mathfrak{o}$ denote the ring of integral functions of x in Σ and $\mathfrak{o}'$ the ring of integral functions of $1/x$ in Σ. If ω is an element of $\mathfrak{o}$, the "exponent" of ω (in symbols exp ω) is defined to be the smallest integer r such that ω/x^r is an element of $\mathfrak{o}'$ (that is $\omega/x^r \subset \mathfrak{o}'$, $\omega/x^{r-1} \not\subset \mathfrak{o}'$). A set of functions $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ is constructed as follows. Let $\lambda_1 = 1$, and let λ_2 be an element in $\mathfrak{o}$ of lowest exponent r_2, such that λ_2 does not satisfy a congruence of the form $\lambda_2 \equiv c\lambda_1 (\mathfrak{o}x)$ where $c \in K$. If $\lambda_1, \lambda_2, \cdots, \lambda_{i-1}$ have been selected, take λ_i to be an element in $\mathfrak{o}$ of lowest exponent which does not satisfy a congruence of the form $\lambda_i \equiv c_1\lambda_1 + \cdots + c_{i-1}\lambda_{i-1} (\mathfrak{o}x)$, $c_j \in K$, $j = 1, \cdots, i-1$. Dedekind and Weber show that this construction leads to a set of $\nu (= [\Sigma : K(x)])$ functions $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ in $\mathfrak{o}$ which have the following properties:

Presented to the Society, September 13, 1943; received by the editors September 10, 1942 and, in revised form, February 17, 1943.

(1) National Research Fellow, 1940–1941.

(a) They are linearly independent over $K(x)$ and therefore form a modular basis for Σ over $K(x)$.

(b) They form a modular basis for $\mathfrak{o}$ over the polynomial ring $K[x]$.

(c) If $\exp \lambda_i = r_i$, then $r_1(=0), 1 \leqq r_2 \leqq \cdots \leqq r_\nu$. If ω is in $\mathfrak{o}$, and if r_h is the last integer in the sequence $r_1, r_2, \cdots, r_\nu$ which is less than or equal to $\exp \omega$, then ω is of the form $\omega = P_1(x)\lambda_1 + \cdots + P_h(x)\lambda_h$, and the degree of $P_i(x)$ (a polynomial) is less than or equal to $(\exp \omega - r_i)$.

(d) The quantities $\lambda_i' = \lambda_i/x^{r_i}$, $i=1, 2, \cdots, \nu$, form a base for $\mathfrak{o}'$ over $K[1/x]$ with the same properties relative to $\mathfrak{o}'$ as the quantities $\lambda_1, \cdots, \lambda_\nu$ have relative to $\mathfrak{o}$.

(e) $\sum_{i=2}^{\nu}(r_i - 1) = p$, the genus of Σ.

Dedekind and Weber refer to such a set of functions as a *normal* base for $\mathfrak{o}$ over $K[x]$.

When we pass to the case of several variables, $x_1, x_2, \cdots, x_r$, and let Σ denote an algebraic extension of $K(x_1, x_2, \cdots, x_r)$, we must consider together with the integral closure $\mathfrak{o}_0$ of the polynomial ring $K[x_1, x_2, \cdots, x_r]$ the integral closures $\mathfrak{o}_i$ of the polynomial rings $K[x_1/x_i, \cdots, x_{i-1}/x_i, 1/x_i, x_{i+1}/x_i, \cdots, x_r/x_i]$, $i=1, 2, \cdots, r$, if we want to generalize the concepts mentioned above. It will be pointed out in §2 that it is possible to define an integer, $\exp \omega$, for every element ω in $\mathfrak{o}_0$ (or in $\mathfrak{o}_i$) with the property that if $\exp \omega = h$ then $\omega/x_i^h \subset \mathfrak{o}_i$, $\omega/x_i^{h-1} \not\subset \mathfrak{o}_i$, $i=1, 2, \cdots, r$. A set of $\nu(=[\Sigma : K(x_1, x_2, \cdots, x_r)])$ elements, $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ in $\mathfrak{o}_0$ which form a modular base for $\mathfrak{o}_0$ over $K[x_1, x_2, \cdots, x_r]$ (and consequently form a base for Σ over $K(x_1, x_2, \cdots, x_r)$) will be referred to as a *DW-base* for $\mathfrak{o}_0$ relative to $K[x_1, x_2, \cdots, x_r]$ if in addition it satisfies the following conditions.

(a) If $\exp \lambda_i = r_i$ then $r_1(=0), 1 \leqq r_2 \leqq r_3 \leqq \cdots \leqq r_\nu$, and if ω is in $\mathfrak{o}_0$ and r_h is the last integer in the sequence $r_1 \leqq r_2 \leqq \cdots \leqq r_\nu$ which is less than or equal to $\exp \omega$, then ω is of the form $\omega = P_1(x)\lambda_1 + P_2(x)\lambda_2 + \cdots + P_h(x)\lambda_h$ where the polynomial P_i is of degree less than or equal to $(\exp \omega - r_i)$.

(b) The quantities $\lambda_i^{(j)} = \lambda_i/x_j^{r_i}$, $j=1, 2, \cdots, r$; $i=1, 2, \cdots, \nu$ have the same properties relative to the ring $\mathfrak{o}_j$ as the elements $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ have relative to $\mathfrak{o}_0$.

In §2 we show that if $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ are the homogeneous coordinates of the general point of an arithmetically normal variety V_r in P_n and if every element of the integrally closed ring $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ depends integrally on the algebraically independent elements $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ then the existence of a DW-base in $\mathfrak{o}_0 = K[\xi_1^*/\xi_0^*, \xi_2^*/\xi_0^*, \cdots, \xi_n^*/\xi_0^*]$ relative to the polynomial ring $K[\xi_1^*/\xi_0^*, \xi_2^*/\xi_0^*, \cdots, \xi_r^*/\xi_0^*]$ is equivalent to the existence of a modular base for $\mathfrak{o}^*$ over $K[\xi_0^*, \xi_1^*, \cdots, \xi_r^*]$ consisting of linearly independent homogeneous elements.

It is well known that such modular bases do not exist in general in rings such as $\mathfrak{o}^*$, so that DW-bases likewise do not always exist.

In this paper we shall prove that if $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ are homogeneous

coordinates along a nonsingular, normal model of a field Σ of degree of transcendency two relative to K, then the existence of a modular base for $\mathfrak{o}^*$ over $K[\xi_0^*, \xi_1^*, \xi_2^*]$ consisting of $\nu(=[\Sigma : K(\xi_1^*/\xi_0^*, \xi_2^*/\xi_1^*)])$ linearly independent homogeneous elements implies that the field Σ is regular. Moreover, if F is a nonsingular model of a regular field Σ and if $\mathfrak{o}_h^*$ is the ring of homogeneous coordinates along a derived normal surface $F^{(h)}$ of F belonging to a sufficiently high character of homogeneity h [9](2), then $\mathfrak{o}_h^*$ has an independent homogeneous modular base relative to a suitably selected set of independent variables.

The proof of the first part of this statement, to the effect that the existence of a base implies regularity, is completely arithmetic in nature. For the proof of the converse, however, we found it necessary to use a lemma of Severi [6] which this author uses in his derivation of the Riemann-Roch theorem for surfaces. The proof of the converse by means of totally arithmetic arguments presents auxiliary problems requiring considerable investigation in themselves. We have not as yet been able to cope with all of the difficulties arising in connection with them.

Throughout this paper we shall use the term *normal* to mean arithmetically normal. The coefficient field K is always assumed to be algebraically closed and of characteristic zero.

1. **Definitions and preliminary remarks.** Let Σ be a field of algebraic functions of degree of transcendency r relative to a subfield K. By a *prime divisor* of Σ we shall mean any $(r-1)$-dimensional valuation of Σ. If $\mathfrak{J}$ is any integral domain in Σ, and if $\mathfrak{P}$ is a prime divisor whose v-ring $\mathfrak{B}(\mathfrak{P})$ contains $\mathfrak{J}$, then $\mathfrak{P}$ is said to be of first or second kind with respect to $\mathfrak{J}$ according as the center(3) of $\mathfrak{P}$ in $\mathfrak{J}$ is or is not $(r-1)$-dimensional.

If $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ are the homogeneous coordinates of the general point of a projective model V_r of Σ, and if $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ is the associated integral domain, then according to a well known theorem of Noether, a nonsingular projective transformation can be found which will insure that the first $r+1$ ξ^*'s, $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ are algebraically independent and that the remaining ones, $\xi_{r+1}^*, \cdots, \xi_n^*$ (and hence the entire ring $\mathfrak{o}^*$), will be *integrally* dependent on $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$. If, then, we specialize $\mathfrak{o}^*$ to the various affine rings $\mathfrak{o}_j$, $\mathfrak{o}_j = K[\xi_0^*/\xi_j^*, \xi_1^*/\xi_j^*, \cdots, \xi_{j-1}^*/\xi_j^*, \xi_{j+1}^*/\xi_j^*, \cdots, \xi_n^*/\xi_j^*]$, $j=0, 1, 2, \cdots, r$, it follows that in each the first r coordinates are algebraically independent and the rest depend integrally on them. It is well known that if $\mathfrak{P}$ is a prime divisor of Σ such that $\mathfrak{B}(\mathfrak{P})$ contains two of the rings $\mathfrak{o}_j$, say $\mathfrak{o}_\alpha$, $\mathfrak{o}_\beta \subset \mathfrak{B}(\mathfrak{P})$, and if $\mathfrak{P}$ is of first kind with respect to $\mathfrak{o}_\alpha$, then $\mathfrak{P}$ is also of the first kind with respect to $\mathfrak{o}_\beta$. Consequently, one may speak of a prime divisor as being of first or second kind relative to a given projective model V_r

(2) Numbers in brackets refer to the bibliography at the end of this paper.

(3) By the *center* of $\mathfrak{P}$ in $\mathfrak{J}$ is meant the prime ideal in $\mathfrak{J}$ consisting of those elements η such that $v(\mathfrak{P}, \eta) > 0$. Throughout this paper we use the notation $v(\mathfrak{P}, x)$ to denote the value of x in the valuation $\mathfrak{P}$.

of Σ. Thus we may associate with a model V_r of Σ a set $\mathfrak{S}(V_r)$ of prime divisors, consisting of all those prime divisors which are of first kind with respect to V_r.

Using the elements of $\mathfrak{S}(V_r)$ as generators we form the abelian group $\mathfrak{G}(V_r)$ which consists of all finite, formal, power products $\mathfrak{P}_1^{\alpha_1}\mathfrak{P}_2^{\alpha_2}\cdots\mathfrak{P}_k^{\alpha_k}$ where $\mathfrak{P}_i \in \mathfrak{S}(V_r)$ and α_i is an integer, $i=1, 2, \cdots, k$. The elements of $\mathfrak{G}(V_r)$ are called divisors. The divisor $\mathfrak{A}=\prod_{i=1}^{k}\mathfrak{P}_i^{\alpha_i}$ is said to be integral if $\alpha_i>0$, $i=1, 2, \cdots, k$.

With every element η in Σ one can associate a uniquely determined divisor $\mathfrak{A}(\eta)$ in $\mathfrak{G}(V_r)$,

$$\mathfrak{A}(\eta) = \prod \mathfrak{P}^{v(\mathfrak{P})}, \qquad v(\eta) = v(\mathfrak{P}, \eta),$$

where the product is extended over all prime divisors of the set $\mathfrak{S}$. (When no ambiguity will result, we shall write η instead of $\mathfrak{A}(\eta)$.) Those divisors which are associated in this way with elements η in Σ form a subgroup $\mathfrak{H}$ of $\mathfrak{G}(V_r)$. Two divisors $\mathfrak{A}$ and $\mathfrak{B}$ are said to be equivalent if $\mathfrak{A}\cdot\mathfrak{B}^{-1}\in\mathfrak{H}$.

With every prime divisor $\mathfrak{P}$ in $\mathfrak{S}$ one can associate a unique, irreducible, $(r-1)$-dimensional subvariety, $V_{r-1}(\mathfrak{P})$, of V_r. With every integral divisor $\mathfrak{A}$ in $\mathfrak{G}(V_r)$, say $\mathfrak{A}=\prod\mathfrak{P}_i^{\alpha_i}$, we associate the effective subvariety

$$V(\mathfrak{A}) = A = \alpha_1 V(\mathfrak{P}_1) + \alpha_2 V(\mathfrak{P}_2) + \cdots + \alpha_k V(\mathfrak{P}_k).$$

The complete system determined by $\mathfrak{A}$ (or A) is, then, the set of all effective subvarieties V ($\mathfrak{A}'$) which are associated with integral divisors $\mathfrak{A}'$ which are equivalent to $\mathfrak{A}$.

In what is to follow, we shall need the following lemma, proved in [4].

LEMMA 1.1 *The algebraically independent quantities* $\xi_1, \xi_2, \cdots, \xi_r$ $(\xi_i=\xi_i^*/\xi_0^*, i=1, 2, \cdots, r)$ *have divisor representations of the form* $\xi_i=\mathfrak{A}_i/\mathfrak{A}_0$, $i=1, 2, \cdots, r$, *where* $\mathfrak{A}_i$ *and* $\mathfrak{A}_0$ *are integral divisors and no prime divisor divides both* $\mathfrak{A}_i$ *and* $\mathfrak{A}_0$.

The following lemma is also needed.

LEMMA 1.2. *If* $\mathfrak{P}$ *is a prime factor of* $\mathfrak{A}_0$, *say* $\mathfrak{A}_0=\mathfrak{P}^{\beta}\mathfrak{A}_0'$ *where* $\mathfrak{A}_0'$ *is prime to* $\mathfrak{P}$, *and if* $f(\xi_1, \xi_2, \cdots, \xi_r)$ *is any polynomial in* $\xi_1, \xi_2, \cdots, \xi_r$ *of degree* h, *then* $v(\mathfrak{P}, f)=-h\beta$. (The notation is that of the previous lemma.)

Proof. Assume first of all that $f=\phi_h$, a homogeneous polynomial in $\xi_1, \xi_2, \cdots, \xi_r$ of degree h, so that $\phi_h/\xi_1^h=\phi_h(1, \xi_2/\xi_1, \cdots, \xi_r/\xi_1)$. This element is in o_1, and since $\mathfrak{P}$ is of first kind, and at finite distance with respect to o_1, the elements $\xi_2/\xi_1, \xi_3/\xi_1, \cdots, \xi_r/\xi_1$ must be algebraically independent modulo $\mathfrak{p}_1$, the center of $\mathfrak{P}$ in o_1. Hence $v(\mathfrak{P}, \phi_h/\xi_1^h)=0$ and $v(\mathfrak{P}, \phi_h)=hv(\mathfrak{P}, \xi_1)=-h\beta$.

If f is an arbitrary polynomial then the lemma follows for f also since $v(\mathfrak{P}, a+b) = \min(v(\mathfrak{P}, a), v(\mathfrak{P}, b))$ if $v(\mathfrak{P}, a) \neq v(\mathfrak{P}, b)$.

It is not difficult to see that if ω is any integral function of $\xi_1, \xi_2, \cdots, \xi_r$ then ω admits a divisor representation of the form

$$\omega = \mathfrak{B}/\mathfrak{A}_0^k$$

where $\mathfrak{B}$ is an integral divisor and k is an integer. In fact, if ω is an integral function of $\xi_1, \xi_2, \cdots, \xi_r$, then ω has non-negative value at any prime divisor at which the values of $\xi_1, \xi_2, \cdots, \xi_r$ are non-negative. Conversely, any element ω in Σ which admits a representation of the form $\omega = \mathfrak{B}/\mathfrak{A}_0^k$ is an integral function of $\xi_1, \xi_2, \cdots, \xi_r$.

2. Exponents of integral functions. We now show that the Dedekind-Weber method of assigning exponents to integral functions can be carried over to the case of functions of several variables almost verbatim. The real significance of these exponents is intimately connected with the notion of arithmetically normal varieties, and we shall treat the subject from this point of view.

Let $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ be the homogeneous coordinates of the general point of a variety V_r in the projective space P_n. It is assumed that V_r is arithmetically normal so that $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ is integrally closed in its quotient field Σ^*. As in the preceding section we assume that $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ are algebraically independent and that $\mathfrak{o}^*$ depends integrally on these $r+1$ elements. We again consider the rings $\mathfrak{o}_0, \mathfrak{o}_1, \cdots, \mathfrak{o}_r$, $\mathfrak{o}_i = K[\xi_0^*/\xi_i^*, \cdots, \xi_{i-1}^*/\xi_i^*, \xi_{i+1}^*/\xi_i^*, \cdots, \xi_n^*/\xi_i^*]$. If λ is any element in $\mathfrak{o}_0$, then λ is a polynomial $f(\xi_1, \xi_2, \cdots, \xi_n)$ $(\xi_i = \xi_i^*/\xi_0^*)$ of degree h, say. Consequently, $\xi_0^{*h} \cdot \lambda = \lambda^*$ is an element of $\mathfrak{o}^*$ which is homogeneous of degree h. It is pointed out in [9] that any element λ^* in $\mathfrak{o}^*$ which is homogeneous of degree h satisfies an equation of the form

$$\lambda^{*\nu} + a_1(\xi_0^*, \cdots, \xi_r^*)\lambda^{*\nu-1} + \cdots + a_\nu(\xi_0^*, \cdots, \xi_r^*) = 0 \tag{2.1}$$

where a_i is a *form* in $\xi_0^*, \xi_1^*, \cdots, \xi_r^*$ of degree $h \cdot i$, $i = 1, 2, \cdots, \nu$ $(\nu = [\Sigma^* : K(\xi_0^*, \xi_1^*, \cdots, \xi_r^*)])$. If we divide equation (2.1) by $\xi_i^{*h\nu}$, $i = 0, 1, \cdots, r$, we find

$$(\lambda^*/\xi_i^{*h})^\nu + (a_1/\xi_i^{*h})(\lambda^*/\xi_i^{*h})^{\nu-1} + \cdots + a_\nu/\xi_i^{*h\nu} = 0.$$

In other words $\lambda^*/\xi_i^{*h} (= \lambda/\xi_i^h)$ is integrally dependent upon $\mathfrak{o}_i$ and hence is in $\mathfrak{o}_i$. We thus see that given any element λ in $\mathfrak{o}_0$ there exist integers h such that $\lambda/\xi_i^h \subset \mathfrak{o}_i$, $i = 1, 2, \cdots, r$. Clearly, if this is true for one integer h then it is true for any integer k greater than h. If for some i, the integer h is such that $\lambda/\xi_i^h \subset \mathfrak{o}_i$, then it is also true that $\lambda/\xi_j^h \subset \mathfrak{o}_j$, for any $j = 1, 2, \cdots, r$. In fact, since λ is in $\mathfrak{o}_0$, we have

$$\lambda = \mathfrak{B}/\mathfrak{A}_0^k \tag{2.2}$$

where $\mathfrak{B}$ is integral, and

$$\lambda/\xi_i^h = (\mathfrak{B}/\mathfrak{A}_0^k)\cdot(\mathfrak{A}_0^k/\mathfrak{A}_i^h) = \mathfrak{B}/\mathfrak{A}_0^{k-h}\cdot\mathfrak{A}_i^h. \tag{2.3}$$

If k is the smallest integer such that (2.2) holds (that is, if $\mathfrak{B}$ is not a multiple of $\mathfrak{A}_0$), then the fact that $\lambda/\xi_i^h \subset \mathfrak{o}_i$ implies that $h \geqq k$, for $\lambda/\xi_i^h \subset \mathfrak{o}_i$, implies that λ/ξ_i^h may be written in the form

$$\lambda/\xi_i^h = \mathfrak{M}/\mathfrak{A}_i^s \tag{2.4}$$

where $\mathfrak{M}$ is integral. Equations (2.3) and (2.4) then imply that $\mathfrak{A}_0^{k-1}\cdot\mathfrak{A}_i^h\cdot\mathfrak{M} = \mathfrak{B}\mathfrak{A}_i^s$ and, since $\mathfrak{A}_i$ and $\mathfrak{A}_0$ are relatively prime, $k \leqq h$ for $\mathfrak{B}$ is not divisible by $\mathfrak{A}_0$. Since we can represent λ in the form $\lambda = \mathfrak{B}'/\mathfrak{A}_0^h$ it follows that $\lambda/\xi_j^h = \mathfrak{B}'/\mathfrak{A}_j^h$, and hence that $\lambda/\xi_j^h \subset \mathfrak{o}_j$.

We therefore define the exponent of the function λ (notation: $\exp\lambda$) to be the smallest integer h such that $\lambda/\xi_i^h \subset \mathfrak{o}_i$. The exponent depends only upon λ and not upon the particular ξ_i which we take to define it. In the case of one independent quantity, ξ_1, this definition is precisely that of Dedekind-Weber.

Several obvious, but important, properties of exponents suggest themselves immediately. We list them below.

(a) $\exp\lambda = h$ *if and only if* $\lambda = \mathfrak{B}/\mathfrak{A}_0^h$, *where* $\mathfrak{B}$ *is an integral divisor not divisible by* $\mathfrak{A}_0$.

In fact, if $\lambda = \mathfrak{B}/\mathfrak{A}_0^h$ then clearly $\exp\lambda \leqq h$, but if $\exp\lambda = h' < h$, then $\lambda = \mathfrak{B}'/\mathfrak{A}_0^{h'}$ as has been shown. This implies that $\mathfrak{B}$ is divisible by $\mathfrak{A}_0$, which is not the case.

(b) *As a corollary to* (a) *we have* $\exp\lambda = 0$ *if and only if* $\lambda \in K$.

(c) *If* η *and* ζ *are two elements of* $\mathfrak{o}_0$, *then* $\exp(\eta+\zeta) \leqq \max(\exp\eta, \exp\zeta)$, *and if* $\exp\eta > \exp\zeta$ *then* $\exp(\eta+\zeta) = \exp\eta$.

If $\exp\eta = h$ and $\exp\zeta = s$, and if $h \geqq s$ then η/ξ_1^h and ζ/ξ_1^h are in $\mathfrak{o}_1$. Hence $\exp(\eta+\zeta) \leqq h$. If $h > s$, then $(\eta+\zeta)/\xi_1^{h-1}$ is not in $\mathfrak{o}_1$ for if it were then we would have $\eta/\xi_1^{h-1} = (\eta+\zeta)/\xi^{h-1} - \zeta/\xi_1^{h-1}$ in $\mathfrak{o}_1$, which is not the case.

(d) *If* $\lambda \in \mathfrak{o}_0$ *and* $\exp\lambda = h$, *then* λ *is a polynomial in* $\xi_1, \xi_2, \cdots, \xi_n$ *of degree less than or equal to* h.

This statement holds in view of the fact that V is a normal model of Σ and hence the system of hyperplane sections on F and all of its multiples are complete [4].

(e) *If* $\lambda = f(\xi_1, \xi_2, \cdots, \xi_n)$ *is a polynomial of degree* h *then* $\exp\lambda \leqq h$.

(f) *If* η *and* ζ *are in* $\mathfrak{o}_0$, *then* $\exp(\eta\cdot\zeta) \leqq \exp\eta + \exp\zeta$. *If* η *is a polynomial in the independent elements* $\xi_1, \xi_2, \cdots, \xi_r$ *only, then* $\exp(\eta\cdot\zeta) = \exp\eta + \exp\zeta$.

The first statement follows directly from (a); the second fellows from (a) in view of Lemma (1.2).

(g) *If* $\mathfrak{A}_0$ *is a prime divisor* $\mathfrak{P}$ *then the equality sign always holds in* (f). *Moreover,* $\exp\lambda = -v(\mathfrak{P}, \lambda)$.

(h) *If* $\lambda^*=\phi(\xi_0^*, \xi_1^*, \cdots, \xi_n^*)$ *is an element of* $\mathfrak{o}^*$ *of degree of homogeneity* h *then* $\exp\lambda$ $(\lambda=\phi/\xi_0^{*h})$ *is less than* h *if and only if* ϕ *is divisible by* ξ_0^* *in* $\mathfrak{o}^*$.

Assertion (h) follows almost immediately from (d) and (e). If $\exp\lambda=k<h$ then, by (d), λ is a polynomial of degree less than or equal to k in $\xi_1, \xi_2, \cdots, \xi_n$, $\lambda=f(\xi_1, \xi_2, \cdots, \xi_n)$, say. Consequently $\psi=\xi_0^{*k}f$ is an element of $\mathfrak{o}^*$, and $\phi=\xi_0^{*h-k}\psi$. On the other hand, if $\phi=\xi_0^{*\sigma}\phi'$, $\sigma>0$, ϕ' in $\mathfrak{o}^*$, then $\lambda=\phi/\xi_0^{*h}=\phi'/\xi_0^{*h-\sigma}=g(\xi_1, \xi_2, \cdots, \xi_n)$ where g is a polynomial of degree $h-\sigma$. By (e) this implies $\exp\lambda\leqq h-\sigma$.

THEOREM 2.1. *If* $\mathfrak{o}^*$ *has an independent modular base over* $K[\xi_0^*, \xi_1^*, \cdots, \xi_r^*]$ *consisting of* ν *homogeneous elements* $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ *where* $\nu=[\Sigma:K(\xi_1, \cdots, \xi_r)]$ *then* $\mathfrak{o}_0$ *(and of course also* $\mathfrak{o}_j$, $j=0, 1, \cdots, r$*) possesses a DW-base over* $K[\xi_1, \xi_2, \cdots, \xi_r]$ *and conversely.*

Proof. Assume that the elements $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ are arranged in the order of increasing degree, so that if λ_i^* is homogeneous of degree r_i we have $r_1\leqq r_2\leqq\cdots\leqq r_\nu$. If ω^* is any homogeneous element of $\mathfrak{o}^*$ of degree h, we have

$$(2.5)\quad \omega^*=P_1^*\lambda_1^*+P_2^*\lambda_2^*+\cdots+P_\nu^*\lambda_\nu^*,\ P_j^*\in K[\xi_0^*, \xi_1^*, \cdots, \xi_r^*].$$

If $P_i^*=\sum P_{ij}^*$, where P_{ij}^* is homogeneous of degree j, it follows immediately by an application of the automorphism $\tau:\xi_0^*\to t\xi_0^*$ of Σ^* over Σ $(\Sigma^*=K(\xi_0^*, \xi_1^*, \cdots, \xi_n^*)$ that $P_{ij}^*\equiv 0$ if $j\neq h-r_i$, and that $P_i^*=P_{ih-r_i}^*$. In other words, in the representation (2.5) of the homogeneous element ω^*, the coefficient P_i^* of λ_i^* is homogeneous of degree $h-r_i$.

It follows from the above remarks that none of the elements λ_i^* is divisible by ξ_0^* in $\mathfrak{o}^*$. In fact, suppose $\lambda_i^*=\xi_0^*\omega^*$. Since ω^* is of degree r_i-1 (we assume for the moment that $i>1$) we have $\omega^*=P_1^*\lambda_1^*+\cdots+P_{i-1}^*\lambda_{i-1}^*$, and hence $\lambda_i^*=Q_1^*\lambda_1^*+\cdots+Q_{i-1}^*\lambda_{i-1}^*$, where $Q_j^*=\xi_0^*P_j^*$. This is in contradiction with the independence of the elements λ_i^*. In particular, λ_1^* must be a constant, which we may take to be 1.

If we let $\lambda_i=\lambda_i^*/\xi_0^{*r_i}$, it follows that λ_i is an element of $\mathfrak{o}_0$ of exponent r_i. If ω is an element of $\mathfrak{o}_0$ of exponent h, then $\xi_0^{*h}\omega$ is homogeneous of degree h, and is in $\mathfrak{o}^*$(4). Hence we have $\xi_0^{*h}\omega=P_1^*\lambda_1^*+P_2^*\lambda_2^*+\cdots+P_\nu^*\lambda_\nu^*$ where P_i^* is homogeneous of degree $h-r_i$. Consequently, we have $\omega=\sum P_i\lambda_i$, where $P_i=P_i^*/\xi_0^{*h-r_i}$. This implies of course that $\exp P_i$ (=degree of P_i) is less than or equal to $h-r_i$.

Since this argument could be carried out with any of the rings $\mathfrak{o}_j$ as well as with $\mathfrak{o}_0$, it follows that the elements $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ have the properties of a DW-base. (See introduction.)

Conversely, if $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ form a DW-base in $\mathfrak{o}_0$ relative to the ring

(4) This assertion follows by property (d) above. In fact, since $\exp\omega=h$, ω is a polynomial of degree h in $\xi_1, \xi_2, \cdots, \xi_n$ and hence $\xi_0^{*h}\omega$ is a form of degree h in $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$.

$K[\xi_1, \xi_2, \cdots, \xi_r]$ and if $\exp \lambda_i = r_i$ it follows immediately that the elements $\lambda_i^* = \xi_0^{*r_i}\lambda_i$ form a modular base for $\mathfrak{o}^*$ over $K[\xi_0^*, \xi_1^*, \cdots, \xi_r^*]$, q.e.d.

3. **Differential divisors and the geometric genus.** We shall henceforth assume that we are dealing with a field Σ of degree of transcendency two over the coefficient field K. We use the same notations as before except that we put $r=2$. We select the quantities $\xi_0^*, \xi_1^*, \xi_2^*$ in such a way that in addition to our previous requirements we also have $[\Sigma : K(\xi_1, \xi_2)] = \nu$, the order of our model F. (We put $\xi_i = \xi_i^*/\xi_0^*$, $i = 1, 2, \cdots, n$.)

Let Δ be any subfield of Σ such that Σ is algebraic over Δ. If $\mathfrak{P}$ is any prime divisor of Σ with valuation ring $\mathfrak{V}$, then $\mathfrak{V}_\Delta = \mathfrak{V} \cap \Delta$ is a valuation ring in Δ which defines a prime divisor $\mathfrak{P}_\Delta$ in Δ. The divisor $\mathfrak{P}_\Delta$ is the valuation which $\mathfrak{P}$ induces in Δ. If $\mathfrak{p}_\Delta$ is the ideal of non-units in $\mathfrak{V}_\Delta$, and $\mathfrak{p}$ is the ideal of non-units in $\mathfrak{V}$, then the extended ideal $\mathfrak{p}_\Delta \cdot \mathfrak{V}$ is a power of $\mathfrak{p}$, say $\mathfrak{p}_\Delta \cdot \mathfrak{V} = \mathfrak{p}^e$, $e \geqq 1$. The divisor $\mathfrak{P}$ is said to be ramified with respect to the field Δ if $e > 1$, unramified if $e = 1$. The integer $e-1$ is called the ramification degree of $\mathfrak{P}$ relative to Δ.

We restrict ourselves to the case in which $\Delta = K(\xi_1, \xi_2)$. We consider the ideal $Z_{(\xi_1, \xi_2)} = \text{g.c.d.}\ (\partial f/\partial z)_{z=\omega}$ [9], where ω is an arbitrary element of $\mathfrak{o}_0$ and $f(z; \xi_1, \xi_2) = N(z-\omega)$. If $\mathfrak{p}_1, \mathfrak{p}_2, \cdots, \mathfrak{p}_s$ are the one-dimensional prime ideals in $\mathfrak{o}_0$ which occur in the decomposition of $Z_{(\xi_1, \xi_2)}$ then we have

$$Z_{(\xi_1, \xi_2)} \approx \mathfrak{p}_1^{\alpha_1} \cdot \mathfrak{p}_2^{\alpha_2} \cdots \mathfrak{p}_s^{\alpha_s} \tag{3.1}$$

where the symbol $\approx$ denotes quasi-equality in the sense of van der Waerden. It is well known that if $\mathfrak{P}$ is ramified with respect to Δ, and if the v-ring $\mathfrak{V}$ of $\mathfrak{P}$ contains $\mathfrak{o}_0$, then $\mathfrak{p}$, the center of $\mathfrak{P}$ in $\mathfrak{o}_0$, is one of the primes $\mathfrak{p}_1, \mathfrak{p}_2, \cdots, \mathfrak{p}_s$ and conversely. Moreover, the ramification degree of $\mathfrak{P}$ relative to Δ is precisely the exponent α assigned to its center in the decomposition (3.1). We conclude therefore that there are only a finite number of prime divisors in the set $\mathfrak{S}(F)$ which are ramified over the field Δ $(= K(\xi_1, \xi_2))$.

The assertions of the preceding paragraph are all consequences of the following lemma.

LEMMA 3.1. *If $\mathfrak{P}$ is any prime divisor of the field Σ (at finite distance relative to the ring $\mathfrak{o}_0$) of ramification degree $e-1$ relative to $K(\xi_1, \xi_2)$, then there is a primitive element θ in $\mathfrak{o}_0$ such that $f'(\theta) \equiv 0\ (\mathfrak{p}^{(e-1)})$, $f'(\theta) \not\equiv 0\ (\mathfrak{p}^{(e)})$, where $\mathfrak{p}$ is the center of $\mathfrak{P}$ in $\mathfrak{o}_0$, and $f'(\theta) = (\partial f/\partial z)_{z=\theta}$. Moreover, it may be assumed that θ has the form $\theta = c_0 + c_1\xi_1 + c_2\xi_2 + \cdots + c_n\xi_n$, $c_i \in K$.*

Proof. The proof is similar to the considerations set forth in [10, paragraphs 11, 12, and 13]. We assume that ξ_1 is transcendental mod $\mathfrak{P}$ so that the residue field of $\mathfrak{P}$ is algebraic over the field $K(\xi_1)$. We let L denote the least normal (Galoisian) extension of $K(\xi_1)$ which contains the residue field of $\mathfrak{P}$. Together with $\mathfrak{o}_0$ we consider the rings $\mathfrak{o}_0' = K(\xi_1)[\xi_2, \xi_3, \cdots, \xi_n]$ and $\mathfrak{O}_0 = L[\xi_2, \xi_3, \cdots, \xi_n]$. In view of the fact that ξ_1 is transcendental mod $\mathfrak{p}$

there is one and only one prime ideal $\mathfrak{p}'$ in $\mathfrak{o}_0'$ such that $\mathfrak{p}' \cap \mathfrak{o}_0 = \mathfrak{p}$. Moreover, if $\mathfrak{J}'$ is the quotient ring of $\mathfrak{o}_0'$ at $\mathfrak{p}'$ and $\mathfrak{J}$ that of $\mathfrak{o}_0$ at $\mathfrak{p}$ then clearly $\mathfrak{J} = \mathfrak{J}'$ and consequently the ramification degree of $\mathfrak{P}$ as a "point" of the one-dimensional variety corresponding to $\mathfrak{o}_0'$ is the same as the ramification degree of $\mathfrak{P}$ as a "curve" of the variety belonging to $\mathfrak{o}_0$. Now if $\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_\Lambda$ are the prime ideals in $\mathfrak{O}_0$ which lie over $\mathfrak{p}'$, that is, $\mathfrak{a}_i \cap \mathfrak{o}_0' = \mathfrak{p}'$, then $\mathfrak{O}_0\mathfrak{p}' = [\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_\Lambda]$. Moreover, it follows immediately from the first part of the proof of Theorem 7 in [10] that if m is any non-negative integer then $\mathfrak{a}_i^m \cap \mathfrak{o}_0' = \mathfrak{p}'^m$(5).

Consider the field $L(\xi_2) \supset K(\xi_1, \xi_2)$, and let P be the divisor induced by $\mathfrak{P}$ in $K(\xi_1, \xi_2)$. The divisor $\mathfrak{A}_i$ in $L \cdot \Sigma = \Lambda$ defined by $\mathfrak{a}_i$ in $\mathfrak{O}_0$ induces a divisor A_i in $L(\xi_2)$ which lies over P in $K(\xi_1, \xi_2)$. Let p, a_i be the centers of P and A_i in their respective v-rings, and let $\mathcal{A}_i$ be the center of $\mathfrak{A}_i$ in its v-ring. If $\mathfrak{R}_i$ is the v-ring of $\mathfrak{A}_i$ then we assert that $a_i \cdot \mathfrak{R}_i = \mathcal{A}_i^e$ where $e-1$ is the ramification degree of $\mathfrak{P}$ relative to $K(\xi_1, \xi_2)$. In other words, the ramification degree of $\mathfrak{A}_i$ relative to $L(\xi_2)$ is the same as that of $\mathfrak{P}$ relative to $K(\xi_1, \xi_2)$. To see this, let R be the v-ring of P in $K(\xi_1, \xi_2)$ and let S_i be that of A_i in $L(\xi_2)$. Now $a_i = pS_i$ and hence $a_i \cdot \mathfrak{R}_i = p \cdot \mathfrak{R}_i$. On the other hand, $\mathfrak{R}_i = \mathfrak{J}' \cdot \mathfrak{R}_i$ and $p \cdot \mathfrak{R}_i = (p \cdot \mathfrak{J}') \cdot \mathfrak{R}_i = \mathcal{A}'^e \cdot \mathfrak{R}_i$ where $\mathcal{A}'$ denotes the center of $\mathfrak{P}$ in $\mathfrak{J}'$. Since $\mathcal{A}' \cdot \mathfrak{R}_i = \mathcal{A}_i$ we have $\mathcal{A}'^e \cdot \mathfrak{R}_i = \mathcal{A}_i^e$. This implies the assertion.

The lemma under consideration is well known in the case of functions of one variable. In fact, if $\theta = c_1\xi_1 + c_2\xi_2 + \cdots + c_n\xi_n$, $c_i \in L$, is a primitive element of Λ over $L(\xi_2)$ then for non-special "constants" c_i, we will have $f'(\theta) \equiv 0 (\mathfrak{a}_i^{e-1})$, $f'(\theta) \not\equiv 0 (\mathfrak{a}_i^e)$. Since the values of c_i that must be avoided are those satisfying certain algebraic relations, we can, in particular, find such quantities in K. For such a choice of the c's, θ will be an element of $\mathfrak{o}_0$ and so also will $f'(\theta)$. This proves the lemma in view of the fact that $\mathfrak{a}_i^m \cap \mathfrak{o}_0 = \mathfrak{p}^{(m)}$, which holds, in particular, for $m = e-1, e$.

The ramification divisor $\mathfrak{Z}_\Delta$ of Σ relative to Δ is defined by the relation

$$\mathfrak{Z}_\Delta = \prod \mathfrak{P}^{e-1}, \qquad \mathfrak{P} \in \mathfrak{S}, \tag{3.2}$$

where $e-1$ is the ramification degree of $\mathfrak{P}$ and the product is extended over all prime divisors of the set $\mathfrak{S}(F)$. $\mathfrak{Z}_\Delta$ may of course contain prime factors which are also factors of $\mathfrak{A}_0$. ($\mathfrak{A}_0$ is the common denominator of ξ_1 and ξ_2 mentioned in Lemma 1.1. It is given by the curve $\xi_0^* = 0$ on F.)

We let $\mathfrak{E}_0$ denote the complementary module of $\mathfrak{o}_0$ relative to $K(\xi_1, \xi_2)$. $\mathfrak{E}_0$ consists of those elements ϵ in Σ such that $S(\epsilon\omega)$ is integral for every ω in $\mathfrak{o}_0$. $S(\epsilon\omega)$ denotes the trace of $\epsilon\omega$ relative to $K(\xi_1, \xi_2)$. The module $\mathfrak{E}_0$ is of

(5) The methods of [10, Theorem 7] are applicable here in view of the fact that as a minimal prime ideal of the integrally closed ring $\mathfrak{o}_0$, $\mathfrak{p}$ defines a simple subvariety of the model V_2 associated with $\mathfrak{o}_0$ and hence $\mathfrak{p}'$ defines a simple subvariety (point) of the model V_1' associated with $\mathfrak{o}_0'$.

paramount importance in the theory of ramified divisors. We prove that $\mathfrak{E}_0$ is the inverse system of the ideal $Z_{(\xi_1,\xi_2)}$. To this end two lemmas are needed, the first of which will be used later in another connection.

LEMMA 3.2. *If $\Delta=K(\xi_1, \xi_2)$ and $\mathfrak{P}$ is any prime divisor of the set $\mathfrak{S}(F)$, and if $\mathfrak{P}_1(=\mathfrak{P}), \mathfrak{P}_2, \cdots, \mathfrak{P}_k$ are the complete set of prime divisors in Σ which induce the same valuation in Δ as does $\mathfrak{P}$, then for any function η such that $v(\mathfrak{P}_i, \eta)>0$, $i=1, 2, \cdots, k$, it is true that $v(\mathfrak{P}, S(\eta))>0$, where the trace is taken relative to Δ.*

Proof. Let P be the common contraction of $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k$ with Δ and let Ω be the least normal extension of Δ which contains Σ. If $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_s$ denote the various valuations of Ω which induce the valuation P on Δ, then each of these induces one of the valuations $\mathfrak{P}_i$ on Σ, $i=1, 2, \cdots, k$. The intersection $\mathcal{B}_1\cap\mathcal{B}_2\cap\cdots\cap\mathcal{B}_s$ where $\mathcal{B}_i$ is the v-ring of $\mathcal{P}_i$ is the integral closure (in Ω) of the v-ring $\mathfrak{B}$ of P in Δ [3]. We denote this intersection by D. The ring D contains s distinct maximal prime ideals $\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_s$ where $\mathfrak{a}_i$ consists of those elements δ in D with the property $v(\mathcal{P}_i, \delta)>0$. By hypothesis, η is contained in the intersection $\mathfrak{a}=[\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_s]$. Since the ideals $\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_s$ are conjugate over Δ [3], the intersection $\mathfrak{a}$ is an invariant ideal, and hence not only η but also all of its conjugates are in $\mathfrak{a}$. Hence their sum $S(\eta)$ is in $\mathfrak{a}$. Since $\mathfrak{a}\cap\mathfrak{B}$ is the ideal of non-units in $\mathfrak{B}$, we conclude that $v(P, S(\eta))>0$, q.e.d.

LEMMA 3.3. *An element ϵ of Σ is an element of the module $\mathfrak{E}_0$ if and only if $v(\mathfrak{P}, \epsilon)\geqq -e+1$ for every prime divisor of the set $\mathfrak{S}(F)$ whose v-ring contains $\mathfrak{o}_0$ ($e-1$ denotes the ramification degree of $\mathfrak{P}$).*

Proof. Let ϵ be an element of Σ with this property. If ϵ is not an element of $\mathfrak{E}_0$ then there is an element ω in $\mathfrak{o}_0$ such that $S(\epsilon\omega)$ is not integral. If we restrict our attention to the group $\mathfrak{G}_\Delta$ of divisors in Δ generated by the prime divisors in Δ of first kind with respect to the rings $K[\xi_1, \xi_2]$, $K[1/\xi_1, \xi_2/\xi_1]$, and $K[\xi_1/\xi_2, 1/\xi_2]$, then $S(\epsilon\omega)$ will have a decomposition of the form

$$S(\epsilon\omega) = \left(\left(\prod_{i=1}^{h} P_i^{\alpha_i}\right)\Big/\left(\prod_{j=1}^{k} Q_j^{\beta_j}\right)\right)A_0^{\sigma}$$

where $\alpha_i, \beta_j>0$, σ is an integer, P_i and Q_j are distinct prime divisors in Δ, and A_0 is that divisor in Δ which corresponds to $\mathfrak{A}_0$ in Σ, that is, $\xi_i=A_i/A_0$, $i=1, 2$.

We fix our attention on one of the divisors Q_i, say Q_1. If q_1 is the center of Q_1 in R ($=K[\xi_1, \xi_2]$), then q_1 is a principal ideal, say $q_1=R\cdot\phi$, where ϕ is a polynomial in ξ_1, ξ_2. We consider the extended ideal $\mathfrak{o}_0q_1$ which has a decomposition of the form $\mathfrak{o}_0q_1=[\mathfrak{p}_1^{(f_1)}, \mathfrak{p}_2^{(f_2)}, \cdots, \mathfrak{p}_s^{(f_s)}]$. By definition, it follows that f_i-1 is the ramification degree of the prime divisor of Σ determined by $\mathfrak{p}_i$.

Now by our hypothesis we have $v(\mathfrak{p}_i, \epsilon\omega\phi) \geqq -f_i+1+f_i$ (since $v(\mathfrak{p}_i, \omega) \geqq 0$), and hence $v(\mathfrak{p}_i, \epsilon\omega\phi) > 0$, $i=1, 2, \cdots, s$. By Lemma 3.2 it follows that $v(Q_1, S(\epsilon\omega\phi)) > 0$, or $v(Q_1, S(\epsilon\omega))+1 > 0$, and hence we have $v(Q_1, S(\epsilon\omega)) \geqq 0$, which is a contradiction. We therefore conclude that ϵ is an element of $\mathfrak{E}_0$.

Conversely, it is well known (see for example [5]) that for any primitive element θ in $\mathfrak{o}_0$ the relation $f'(\theta)\cdot\mathfrak{E}_0 \subset \mathfrak{o}_0$, where $f(z)=N(z-\theta)$, holds. Moreover, by Lemma 3.1 it follows that for any prime divisor $\mathfrak{P}$ with center $\mathfrak{p}$ in $\mathfrak{o}_0$ and ramification degree $e-1$ there exists a primitive element θ in $\mathfrak{o}_0$ such that $f'(\theta) \equiv 0$ $(\mathfrak{p}^{(e-1)})$, $f'(\theta) \not\equiv 0$ $(\mathfrak{p}^{(e)})$. From this it follows that for any ϵ in $\mathfrak{E}_0$ the inequality $v(\mathfrak{P}, \epsilon) \geqq -e+1$ holds. This proves the lemma.

This characterization of $\mathfrak{E}_0$ in terms of the ramified divisors at finite distance relative to $\mathfrak{o}_0$ (that is, having v-rings which contain $\mathfrak{o}_0$) is obviously equivalent to the statement that $\mathfrak{E}_0$ is the inverse system of $Z_{(\xi_1,\xi_2)}$ in Σ, in view of the relation (3.1). Making use of this fact, it is not difficult to see that every element ϵ of $\mathfrak{E}_0$ admits a divisor representation of the form

$$\epsilon = (\mathfrak{D}/\mathfrak{Z}_\Delta)\mathfrak{A}_0^r \tag{3.3}$$

where $\mathfrak{D}$ is an integral divisor and r is an integer which may be positive, negative, or zero(6). Conversely, any element of Σ which admits a representation such as (3.3) is in $\mathfrak{E}_0$.

An integral divisor $\mathfrak{D}$ equivalent to the divisor $\mathfrak{Z}_\Delta\mathfrak{A}_0^{-3}$ (in symbols $\mathfrak{D} \sim \mathfrak{Z}_\Delta\mathfrak{A}_0^{-3}$) is called a differential divisor of first kind. To every such divisor $\mathfrak{D}$ there corresponds an element δ, necessarily in $\mathfrak{E}_0$, which expresses the equivalence, namely,

$$\delta = \mathfrak{D}\mathfrak{A}_0^3/\mathfrak{Z}_\Delta. \tag{3.4}$$

The geometric genus p_g of the surface F (a normal nonsingular model) is defined as the number of linearly independent differential divisors of first kind, that is the number of linearly independent elements δ in $\mathfrak{E}_0$ of the form (3.4). A curve K on F determined by a differential divisor of first kind is called a canonical curve, and the complete system $|K|$ is called the pure canonical system on F. The effective dimension of $|K|$ is p_g-1.

---

(6) Equation (3.3) follows immediately if none of the prime factors of $\mathfrak{A}_0$ is ramified. Otherwise, the proof is as follows. If $\mathfrak{P}$ is any prime factor of $\mathfrak{A}_0$, say $\mathfrak{A}_0 = \mathfrak{P}^\nu \cdot \mathfrak{A}_0'$, where $\mathfrak{A}_0'$ is prime to $\mathfrak{P}$, then $e-1$ is the ramification degree of $\mathfrak{P}$. Since $\mathfrak{P}$ is at finite distance relative to the ring $\mathfrak{o}_1$ $(=K[1/\xi_1, \xi_2/\xi_1, \cdots, \xi_n/\xi_1])$ it follows that there exists an element θ_1 in $\mathfrak{o}_1$ of exponent one, such that $f'(\theta_1) \equiv 0$ $(\mathfrak{p}^{(e-1)})$, $f'(\theta_1) \not\equiv 0$ $(\mathfrak{p}^{(e)})$ where $\mathfrak{p}$ is the center of $\mathfrak{P}$ in $\mathfrak{o}_1$. Since θ_1 is of exponent one it follows that $\theta = \xi_1\theta_1$ is in $\mathfrak{o}_0$. A simple computation leads to the conclusion that $v(\mathfrak{P}, f'(\theta)) = -e(\nu-1)+e-1$, so that in the representation $f'(\theta) = \mathfrak{Q}\mathfrak{Z}_\Delta/\mathfrak{A}_0^{\nu-1}$ the divisor $\mathfrak{Q}$ does not have $\mathfrak{P}$ as a factor. Since $\epsilon \cdot f'(\theta) = \mathfrak{M}/\mathfrak{A}_0^b$, where $\mathfrak{M}$ is integral, we have $\epsilon = (\mathfrak{M}/\mathfrak{Q}\mathfrak{Z}_\Delta)\mathfrak{A}_0^{\nu-b-1}$. Applying this process to each of the factors of $\mathfrak{A}_0$ in turn, we find that $\epsilon = (\mathfrak{N}/\mathfrak{U}\mathfrak{Z}_\Delta)\mathfrak{A}_0^r$ where $\mathfrak{U}$ has no factors in common with $\mathfrak{A}_0$. It then follows by Lemma 3.3 that $\mathfrak{N}$ contains $\mathfrak{U}$ as a factor.

As it stands, our definition of differential divisors of first kind and of the geometric genus p_g is but a formal generalization of similar definitions given by Dedekind and Weber in [2]. However, it is not difficult to see that if one considers a generic projection of our normal nonsingular model F into a surface in S_3 given by $f(\xi_1, \xi_2, \theta)=0$ where $\theta=c_0+c_1\xi_1+\cdots+c_n\xi_n$ then the elements $\delta f'(\theta)$ will be adjoint polynomials, $\phi_{\nu-4}$ of degree $\nu-4$, to the surface $f(\xi_1, \xi_2, \theta)=0$. Hence since $\delta=\phi_{\nu-4}/f'(\theta)$ the p_g independent elements δ determine the p_g independent double integrals of first kind attached to F.

4. **The arithmetic genus.** Let $x_0, x_1, \cdots, x_n$ be indeterminates and let P denote the prime H-ideal in $K[x_0, x_1, \cdots, x_n]$ which determines a normal, nonsingular model F of Σ. In other words, P is a prime ideal in $K[x_0, x_1, \cdots, x_n]$ such that $K[x_0, x_1, \cdots, x_n]/P \cong_0 {}^*$ and $x_i \to \xi_i^*$ under the isomorphism. Van der Waerden has shown [7] that if $\chi(P, h)$ denotes the number of homogeneous elements of degree h in $K[x_0, x_1, \cdots, x_n]$ which are linearly independent (over K) mod P, then for large values of h (that is, for $h>h_0$, where h_0 is a fixed integer depending on P) $\chi(P, h)$ is given by

$$\chi(P, h) = a_0C_{h,2} + a_1C_{h,1} + a_2 \tag{4.1}$$

where a_0, a_1, and a_2 are integers independent of h. Van der Waerden showed further that $a_0=\nu$, the order of the surface F.

Formula (4.1) can also be derived by means of the Riemann-Roch theorem for surfaces, with the help of Noether's formulae for the genus and order of a composite curve [8]. The theorem of Riemann-Roch states that if $|D|$ is a complete linear system on a surface F then its dimension r is given by the formula $r=n-\pi+p_a+1-i+s$ where n, π, and i are respectively the degree, genus, and index of speciality of $|D|$, and p_a is the arithmetic genus of F. The non-negative integer s is called the superabundance of the system, and the difference $s-i$ is called the irregularity of $|D|$. The system $|D|$ is said to be regular if $s=i=0$. It has been shown by Severi, that if $|C|$ is a nonsingular system of dimension $r\geqq 3$, then for large values of h the system $\begin{vmatrix} C_h \\ C_h \end{vmatrix} = |hC|$ is a regular system. Applying the Riemann-Roch theorem to $|hC|$, and denoting its dimension, degree and genus by ρ_h, ν_h and π_h respectively, we have

$$\rho_h = \nu_h - \pi_h + p_a + 1. \tag{4.2}$$

If F is a normal, nonsingular model of the field Σ, and $|C|$ is its system of hyperplane sections, then $|C|$ is nonsingular, and the complete system $|hC|$ is cut out on F by the hypersurfaces of order h in its ambient space [4]. Hence in this case, $\chi(P, h)=\rho_h+1$. A straightforward application of Noether's formulae to formula (4.2) yields

$$\chi(P, h) = (h(h-1)/2)\nu + (\nu - \pi + 1)h + p_a + 1, \tag{4.3}$$

where ν is the order of F and π is the genus of $|C|$ (7).

As yet, no purely arithmetic proof of formula (4.2) is available. For the purposes of this paper we shall regard formula (4.3) as the definition of p_a(8).

5. **The existence of a DW-base implies $p_g = p_a$.** Let $\mathfrak{o}^* = K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$ where $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ are the homogeneous coordinates along a normal non-singular model F of the field Σ. We make the following assumptions:

(a) $\xi_0^*, \xi_1^*, \xi_2^*$ are algebraically independent and every element of $\mathfrak{o}^*$ depends integrally on them.

(b) There exist ν (=order of F) homogeneous elements $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ in $\mathfrak{o}^*$ linearly independent over $K[\xi_0^*, \xi_1^*, \xi_2^*]$ with the property that if ω^* is any element in $\mathfrak{o}^*$ then $\omega^* = P_1^*\lambda_1^* + P_2^*\lambda_2^* + \cdots + P_\nu^*\lambda_\nu^*$, where P_i^* is a polynomial in $\xi_0^*, \xi_1^*, \xi_2^*$. We let r_i be the degree of λ_i^*, and we assume as we did in Theorem 2.1 that $r_1 = 0$, $1 \leqq r_2 \leqq r_3 \leqq \cdots \leqq r_\nu$.

In the course of the proof of Theorem 2.1 we saw that under the above assumptions we could assert in addition that if ω^* is a homogeneous element of $\mathfrak{o}^*$ of degree h, and if, say, r_s is the last integer in the sequence $r_1, r_2, \cdots, r_\nu$ such that $h \geqq r_s$ (that is, $h < r_{s+1}$ if λ_{s+1}^* is defined), then $\omega^* = P_1^*\lambda_1^* + P_2^*\lambda_2^* + \cdots + P_s^*\lambda_s^*$ where P_i^* is a homogeneous form in $\xi_0^*, \xi_1^*, \xi_2^*$ of degree $h - r_i$.

The function $\chi(P, h)$ mentioned in the preceding section gives the number of homogeneous elements in $\mathfrak{o}^*$ of degree h which are linearly independent over K. If in particular $h > r_\nu$ we know that any homogeneous element of degree h can be written in the form $P_1^*\lambda_1^* + P_2^*\lambda_2^* + \cdots + P_\nu^*\lambda_\nu^*$ where P_i^* is homogeneous of degree $h - r_i$. Since $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ are linearly independent over $K[\xi_0^*, \xi_1^*, \xi_2^*]$ it follows that a linearly independent K-base for the homogeneous elements of degree h in $\mathfrak{o}^*$ can be obtained by taking together all of the elements of the ν sets of elements of the form $a_i^{(j)}\lambda_i^*$ where $a_i^{(j)}$ runs over a linearly independent base for the homogeneous polynomials in $\xi_0^*, \xi_1^*, \xi_2^*$ of degree $h - r_i$. There are, then, $(1/2)(h - r_i + 1)(h - r_i + 2)$ such elements in the ith set $\{a_i^{(j)}\lambda_i^*\}$ and hence we have

$$\chi(P, h) = \sum_{i=1}^{\nu} \frac{(h - r_i + 1)(h - r_i + 2)}{2}. \tag{5.1}$$

A simple computation (taking into account that $r_1 = 0$) yields

(7) For a complete discussion of these questions see [8, chapter 4].

(8) It seems to us that from the arithmetic point of view, formula (4.3) is a very natural way in which to introduce the arithmetic genus. The question might be raised as to the necessity of requiring F to be *nonsingular* as well as normal in order that formula (4.3) hold. The cone projecting an elliptic cubic from a point not in the plane of the cubic furnishes an example of a surface in S_3 which is arithmetically normal but on which the complete system $|C_h|$ cut out by the surfaces of order h in S_3 has a superabundance $s = +1$ for every h.

$$\chi(P, h) = \frac{h(h-1)}{2}\nu + \left(\nu - \sum_{i=2}^{\nu}(r_i - 1) + 1\right)h + \sum_{i=2}^{\nu}\frac{(r_i-1)(r_i-2)}{2} + 1. \tag{5.2}$$

Combining (5.1) and (4.3) we conclude that

$$p_a = \sum_{i=2}^{\nu}\frac{(r_i-1)(r_i-2)}{2}. \tag{5.3}$$

The proof that $p_g = p_a$ is completed by showing that p_g can also be expressed in terms of the integers r_i, and that the expression one thereby obtains is $p_g = (1/2)\sum_{i=2}^{\nu}(r_i-1)(r_i-2)$. This is somewhat more difficult.

As was pointed out in Theorem 2.1, if $\lambda_i = \lambda_i^*/\xi_0^{*r_i}$ then $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ is a DW-base in $\mathfrak{o}_0$ relative to the ring $K[\xi_1, \xi_2]$. It is well known that there exist ν elements $\epsilon_1, \epsilon_2, \cdots, \epsilon_\nu$ in Σ satisfying the equations $S(\epsilon_i\lambda_j) = \delta_{ij}$ ($\delta_{ii}=1$, $\delta_{ij}=0$, $i \neq j$). These elements ϵ are all in $\mathfrak{E}_0$ (the complementary modulo of $\mathfrak{o}_0$) and, in fact, form a modular base for $\mathfrak{E}_0$ over $K[\xi_1, \xi_2]$, called the complementary base to $\lambda_1, \lambda_2, \cdots, \lambda_\nu$. The elements $\epsilon_1, \epsilon_2, \cdots, \epsilon_\nu$ are of course linearly independent over $K(\xi_1, \xi_2)$. If ϵ is any element of $\mathfrak{E}_0$, then we have $\epsilon = a_1\epsilon_1 + a_2\epsilon_2 + \cdots + a_\nu\epsilon_\nu$, $a_i \in K[\xi_1, \xi_2]$, and it follows from the equation $S(\epsilon_i\lambda_j) = \delta_{ij}$ that the coefficient a_i is equal to $S(\epsilon\lambda_i)$.

We assume, in particular, that ϵ is of the form $\mathfrak{D}\mathfrak{A}_0^3/\mathfrak{Z}_\Delta$ and we assert that if $\mathfrak{P}$ is any factor of $\mathfrak{A}_0$, then $v(\mathfrak{P}, \xi_1^2\epsilon) > 0$. This is obvious if $\mathfrak{P}$ is not a factor of $\mathfrak{Z}_\Delta$. If, on the other hand, $\mathfrak{Z}_\Delta = \mathfrak{P}^\beta\mathfrak{Z}'$ ($\mathfrak{Z}'$ prime to $\mathfrak{P}$), $\beta > 0$, then since $\mathfrak{P}$ induces the divisor P in $K(\xi_1, \xi_2)$ at which $(1/\xi_1)$ has positive value, it follows from the definitions of $\mathfrak{Z}_\Delta$ and $\mathfrak{A}_0$ that $v(\mathfrak{P}, 1/\xi_1) = \beta + 1$, and that $\mathfrak{A}_0 = \mathfrak{P}^{\beta+1}\mathfrak{A}_0'$. We conclude that $v(\mathfrak{P}, \epsilon) \geqq 3(\beta+1) - \beta = 2\beta + 3$ and hence that $v(\mathfrak{P}, \xi_1^2\epsilon) \geqq 2\beta + 3 - 2(\beta+1) = 1$.

Writing ϵ in the form $\epsilon = \sum a_i\epsilon_i$ we have

$$a_i = S(\epsilon\lambda_i), \qquad a_i/\xi_1^{r_i-2} = S(\xi_1^2\epsilon \cdot \lambda_i/\xi_1^{r_i}). \tag{5.4}$$

Since $\lambda_i/\xi_1^{r_i}$ is an element of $\mathfrak{o}_1$ ($\mathfrak{o}_1 = K[\xi_0^*/\xi_1^*, \xi_2^*/\xi_1^*, \cdots, \xi_n^*/\xi_1^*]$) its value at any of the prime factors of $\mathfrak{A}_0$ is non-negative, and hence the value of $\xi_1^2\epsilon\lambda_i/\xi_1^{r_i}$ is positive at any such factor. It follows by Lemma 3.2 that $v(\mathfrak{P}, S(\xi_1^2\epsilon\lambda_i/\xi_1^{r_i})) > 0$, and hence by (5.4), $v(\mathfrak{P}, a_i/\xi_1^{r_i-2}) > 0$ for any such prime factor $\mathfrak{P}$. By Lemma 1.2, the inequality $v(\mathfrak{P}, a_i/\xi_1^{i-2}) > 0$ implies that if $r_i \leqq 2$ then a_i is identically zero, and if $r_i > 2$ the degree of a_i is less than $r_i - 2$.

We prove the converse of this result. That is, if $\epsilon = \sum a_i\epsilon_i$ is an element of $\mathfrak{C}_0$ such that $a_i \equiv 0$ if $r_i \leqq 2$, and for those values of i for which $r_i > 2$ the degree

of a_i is less than r_i-2 then the element ϵ admits a divisor decomposition of the form $\epsilon=\mathfrak{D}\mathfrak{A}_0^3/\mathfrak{Z}_\Delta$. We consider the set of elements λ_1', λ_2', $\cdots$, λ_ν' where $\lambda_i'=\lambda_i/\xi_1^{r_i}$. As we have seen, these elements form a DW-base for the ring $\mathfrak{o}_1$. It is well known that the base ϵ_1', ϵ_2', $\cdots$, ϵ_ν' complementary to $\lambda_1', \lambda_2', \cdots, \lambda_\nu'$ is related to the base $\epsilon_1, \epsilon_2, \cdots, \epsilon_\nu$ by the equations $\epsilon_i'=\xi_1^{r_i}\epsilon_i$. These elements form an independent modular base over $K[1/\xi_1, \xi_2/\xi_1]$ for the complementary module $\mathfrak{E}_1$ of $\mathfrak{o}_1$ relative to the ring $K[1/\xi_1, \xi_2/\xi_1]$.

Let $\mathfrak{P}$ be any factor of $\mathfrak{A}_0$, say $\mathfrak{A}_0=\mathfrak{P}^e\mathfrak{A}_0'$ where $\mathfrak{A}_0'$ is prime to $\mathfrak{P}$. If $e>1$, we have $\mathfrak{P}^{e-1}$ as a factor of $\mathfrak{Z}_\Delta$. Since $\mathfrak{P}$ is a divisor at finite distance in the coordinate system which gives rise to $\mathfrak{o}_1$, and since ϵ_i' is in $\mathfrak{E}_1$ it follows that $v(\mathfrak{P}, \epsilon_i')\geqq -e+1$. Hence we see that $v(\mathfrak{P}, (1/\xi_1)\epsilon_i')\geqq e+(-e+1)=1$. We thus have $v(\mathfrak{P}, \xi_1^{r_i-1}\epsilon_i)>0$ and $v(\mathfrak{P}, \epsilon_i)>e(r_i-1)$. By our assumptions concerning the coefficients a_i we have $v(\mathfrak{P}, a_i)\geqq -e(r_i-3)$, and hence we conclude that $v(\mathfrak{P}, a_i\epsilon_i)>2e$.

The element $a_i\epsilon_i$ is in $\mathfrak{E}_0$, and if a_i is not identically zero we can write

$$a_i\epsilon_i = (\mathfrak{M}/\mathfrak{Z}_\Delta)\mathfrak{P}_1^{s_1}\cdot\mathfrak{P}_2^{s_2}\cdots\mathfrak{P}_k^{s_k}$$

where $\mathfrak{M}$ is an integral divisor and $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k$ are the factors of $\mathfrak{A}_0$. If (e_j-1) is the ramification degree of $\mathfrak{P}_j$ then by our above results we have $s_1-(e_j-1)>2e_j$, and hence $s_1\geqq 3e_j$. Hence, since $\mathfrak{A}_0=\prod_j\mathfrak{P}_j^{e_j}$, we conclude that the element $a_i\epsilon_i$ is of the form

$$a_i\epsilon_i = \mathfrak{A}_0^3\mathfrak{N}/\mathfrak{Z}_\Delta.$$

The integral divisor $\mathfrak{N}$ is therefore a differential divisor of first kind. The number of independent elements of the form $a_i\epsilon_i$ in $\mathfrak{E}_0$ which give rise to differential divisors of first kind is therefore zero if $r_i\leqq 2$ and $(r_i-1)(r_i-2)/2$ (=number of independent a_i of degree less than r_i-2) if $r_i>2$. Since the ϵ_i form a base for $\mathfrak{E}_0$, we conclude that

$$p_g = \sum_{r_i\geq 3}\frac{(r_i-1)(r_i-2)}{2} = \sum_{i=2}^{\nu}\frac{(r_i-1)(r_i-2)}{2}. \tag{5.5}$$

This completes the proof that the existence of a base implies $p_g=p_a$.

6. **A lemma of Castelnuovo.** This section is devoted to establishing an important lemma due originally to Castelnuovo [1]. We consider a normal, nonsingular model F of Σ, and we use the same notations as before. In particular, we consider the various complete systems $|C_h|=|hC|$ cut out on F by the hypersurfaces of order h of its ambient space P_n.

The system $|C_h|$ cuts out a linear series g_{r_h} on a generic curve C of the system of hyperplane sections $|C|$. (In what sense C is generic will be specified in a moment.) The series g_{r_h} has a certain index of speciality i_h and a certain deficiency δ_h. Clearly, if h is large we will have $i_h=0$. Moreover, since C is

generic in $|C|$ it is nonsingular and hence locally normal. This implies that $\delta_h=0$ for large values of h [12]. The lemma of this section asserts the following equality(9)

$$p_a = \sum_{h=1}^{\infty} (i_h - \delta_h). \tag{6.1}$$

Proof. We subject the homogeneous coordinates $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ to a projective transformation $\xi_i=\sum a_{ij}\xi_j^*$, $a_{ij}\in K$. For non-special values of the a_{ij} the property of integral dependence of $\mathfrak{o}^*$ upon the first three coordinates will not be lost. Moreover, since the system $|C|$ is obviously not composite with a pencil, it follows by the theorem of Bertini-Zariski [11] that for non-special values of the constants a_{ij} the curve defined on F by the equation $\xi_2=0$ will be irreducible and nonsingular. In fact, it is easily seen from Zariski's consideration in [11] that for non-special constants a_{ij} the ideal $\mathfrak{o}^*\xi_2$ will be a prime H-ideal in $\mathfrak{o}^*$. (This of course involves the fact that $\mathfrak{o}^*$ is integrally closed.) It is in this sense that we speak of the generic curve in $|C|$. We fix a set of constants a_{ij} so that these requirements are satisfied. The characters i_h and δ_h are understood to refer to the series $g_{\nu h}$ cut out on $\xi_2=0$ by $|C_h|$.

As before we let $\chi(P, h)$ be the number of independent (over K) homogeneous elements of degree h in $\mathfrak{o}^*$, and we let μ_h be the number of homogeneous elements of degree h which are independent (over K) modulo $\mathfrak{o}^*\xi_2$. It is not difficult to see that $\chi(P, h)=\mu_h+\mu_{h-1}+\cdots+\mu_0$. In fact, let $\theta_{1i}, \theta_{2i}, \cdots, \theta_{\mu_i i}$ be a set of μ_i homogeneous elements in $\mathfrak{o}^*$ of degree i which are independent modulo $\mathfrak{o}^*\xi_2$. The $\sum_{i=0}^{h}\mu_i$ elements $\{\cdots, \xi_2^{h-i}\theta_{1i}, \xi_2^{h-i}\theta_{2i}, \cdots, \xi_2^{h-i}\theta_{\mu_i i}, \cdots\}$ are obviously independent, and every homogeneous element of degree h in $\mathfrak{o}^*$ is a linear combination of these elements.

By the Riemann-Roch theorem (for curves), we have $\mu_k=k\nu-\pi+i_k-\delta_k+1$, $k=1, 2, \cdots, h$, where π is the genus of $\xi_2=0$, and since, clearly, $\mu_0=1$ we have

$$\chi(P, h) = 1 + \sum_{k=1}^{h} \{k\nu - \pi + i_k - \delta_k + 1\} \tag{6.2}$$

or on rearranging (6.2) we have

$$\chi(P, h) = \frac{h(h-1)}{2}\nu + (\nu - \pi + 1)h + \sum_{k=1}^{h} (i_k - \delta_k) + 1. \tag{6.3}$$

Since $\sum_{k=1}^{h}(i_k-\delta_k)$ remains constant for sufficiently large h we see, if we compare (6.3) and (4.3), that $p_a=\sum_{k=1}^{\infty}(i_k-\delta_k)$, q.e.d.

(9) The original lemma of Castelnuovo [1] asserts that if one is given a system $|\Gamma|$, then for high values of h the superabundance of $|h\Gamma|$ has a constant value s, and $s+p_a=\sum(i_k-\delta_k)$.

7. **On the indices i_h.** We use the same notations and assumptions as in the preceding section. The index of speciality, j_h, of the system $|C_h|$ is defined as the dimension, increased by 1, of the system $|K-C_h|$ where $|K|$ is the canonical system on F. If $|K-C_h|$ does not exist we put $j_h=0$. We observe first of all that $j_h=0$ if h is sufficiently large.

The lemma of Severi [6] mentioned in the introduction is to the effect that *if $h>\nu-4$ then the canonical system $|K|$, if it exists (that is, if $p_g>0$), cuts out a complete series on a generic curve C_h of $|C_h|$. If $p_g=0$, the characteristic series*[10] *of $|C_h|$ is non-special.* Since, in any case, the series cut out on $|C_h|$ by $|K|$ is totally contained in the series difference of the canonical series on C_h and the characteristic series on C_h, it follows, by Severi's lemma, that for $h>\nu-4$ the complete series difference between the canonical and characteristic series on C_h is cut out by the canonical system $|K|$. If in addition h_0 $(>\nu-4)$ is taken large enough so that $j_{h_0}=0$, then for all $h\geqq h_0$ we have that the index of speciality of the characteristic series of C_h is precisely p_g.

The formulas of Noether [8] yield the following two formulas immediately:

$$\nu_h = h^2\nu_1, \qquad \pi_h = h\pi_1 + (h(h-1)/2)\nu_1 - (h-1) \tag{7.1}$$

where ν_h and π_h denote the degree and genus of $|C_h|$. From these one easily sees that if h is sufficiently large, the inequality $\nu_h>\pi_h-1$ will hold.

Let ρ be an integer greater than or equal to h_0 and in addition large enough so that for $h\geqq\rho$, $\nu_h>\pi_h-1$. We pass from F to the derived normal surface F_ρ of F by referring the curves of $|C_\rho|$ to the sections of F_ρ cut cut by the hyperplanes of its ambient space P_N. Since F is nonsingular, so is F_ρ [9]. If i_h is the index of speciality of the series cut out on a generic C_ρ by the system $|C_{h\rho}|$, then

$$i_1 = p_g, \; i_2 = i_3 = \cdots = 0. \tag{7.2}$$

(The condition $i_2=i_3=\cdots=0$ is implied by the fact that $k\nu_\rho>2\pi_\rho-2$ for $k=2, 3, \cdots$.) A model F of Σ will be called *non-special* if it is normal, nonsingular, and if the speciality indices on its generic hyperplane section satisfy equations (7.2).

We return to the notation of §6, but we now assume that the model F considered there is non-special. Combining (6.1) and (7.2) we find

$$\sum_{h=1}^{\infty} \delta_h = p_g - p_a. \tag{7.3}$$

It follows from (7.3) that if, in particular, F is regular then $\delta_h=0$,

[10] By the characteristic series of $|C_h|$ on C_h we mean, of course, the series cut out on the curve C_h by the complete system $|C_h|$.

$h=1, 2, \cdots$. In other words, if F is a non-special model of a regular field, then the systems $|C_h|$ cut out complete series on the general hyperplane section C of F. This implies that the general hyperplane section of F is arithmetically normal [4].

8. **Construction of an independent base.** Let $\mathfrak{o}^*=K[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$, where $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ are homogeneous coordinates along a *non-special* model F of the field Σ. We assume that the quantities ξ_0^*, ξ_1^* and ξ_2^* are selected so that they are algebraically independent, and so that every element in $\mathfrak{o}^*$ depends integrally on them. The only other restriction which we place on the elements ξ_0^*, ξ_1^*, ξ_2^* is that the net N: $c_0\xi_0^*+c_1\xi_1^*+c_2\xi_2^*=0$ (the c's are parameters which vary over K) be sufficiently general in $|C|$ so that the general curve of N will satisfy the condition of the preceding section, that is, we require that the general curve of N be arithmetically normal. We let $R^*=K[\xi_0^*, \xi_1^*, \xi_2^*]$ and we shall prove that $\mathfrak{o}^*$ has an independent modular base over R^*.

We apply a nonsingular linear transformation $\eta_i=a_{i0}\xi_0^*+a_{i1}\xi_1^*+a_{i2}\xi_2^*$, $a_{ij}\in K$, $i=0, 1, 2$, if necessary, in order to insure that (a) the three ideals $\mathfrak{o}^*\eta_0$, $\mathfrak{o}^*\eta_1$, $\mathfrak{o}^*\eta_2$ will be prime in $\mathfrak{o}^*$, (b) that these curves will intersect pair by pair in ν distinct points of the surface F, each point of intersection being a simple intersection, (c) each of these curves will be arithmetically normal. Since ξ_0^*, ξ_1^*, ξ_2^* are algebraically independent, it follows that the net N is not composite with a pencil, and hence by [11] condition (a) is satisfied for a general choice of the constants $a_{ij}\in K$. It is easily seen that if the coefficients a_{ij} are sufficiently general, conditions (b) and (c) will automatically be satisfied. We point out that since $\mathfrak{o}^*$ depends integrally on η_0^*, η_1^*, η_2^*, the ideal $\mathfrak{o}^*(\eta_0^*, \eta_1^*, \eta_2^*)$ is projectively irrelevant, and hence none of the ν intersection points of $\eta_1=0$ and $\eta_2=0$ can be on the curve $\eta_0=0$.

It should be observed that the transformation from ξ_0^*, ξ_1^*, ξ_2^* to η_0, η_1, η_2 does not change the ring R^*, and hence we may assume without loss of generality that the particular elements ξ_0^*, ξ_1^*, ξ_2^* have all of the special properties which we have assigned to η_0, η_1, η_2.

We pass to the ring $\mathfrak{o}_0=K[\xi_1, \xi_2, \cdots, \xi_n]$ where $\xi_i=\xi_i^*/\xi_0^*$ and we observe that if $\mathfrak{A}=\mathfrak{o}_0(\xi_1, \xi_2)$ then the K-module $\mathfrak{o}_0/\mathfrak{A}$ is of rank ν, the order of F. In fact if $\mathfrak{A}=[\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_s]$ is the primary decomposition of $\mathfrak{A}$ ($\mathfrak{q}_i$ is a zero-dimensional ideal) then by assumption (b) we must have $s=\nu$ and $\mathfrak{q}_i=\mathfrak{p}_i$, $i=1, 2, \cdots, \nu$, where $\mathfrak{p}_i$ is a prime zero-dimensional ideal. It follows immediately that $\mathfrak{A}^*=\mathfrak{o}^*(\xi_1^*, \xi_2^*)$ admits a primary decomposition of the form $\mathfrak{A}^*=[\mathfrak{p}_1^*, \mathfrak{p}_2^*, \cdots, \mathfrak{p}_\nu^*, \mathfrak{q}]$, where $\mathfrak{p}_i^*$ is the homogeneous prime in $\mathfrak{o}^*$ corresponding to the prime $\mathfrak{p}_i$ in $\mathfrak{o}_0$. As an ideal in $\mathfrak{o}^*$, $\mathfrak{p}_i^*$ is one-dimensional, but as a homogeneous ideal its dimension is counted as zero([11]). The ideal $\mathfrak{q}$ is a

([11]) For a complete discussion of H-ideals and their related affine ideals, see the first two sections of van der Waerden [7].

possible embedded component of $\mathfrak{A}^*$ which is necessarily projectively irrelevant, that is, its associated prime in $\mathfrak{o}^*$ is the ideal $\mathfrak{o}^*(\xi_0^*, \xi_1^*, \cdots, \xi_n^*)$. The central point in preparing for the construction of a base is the following lemma.

LEMMA 8.1. *The fact that $\xi_2^*=0$ is an arithmetically normal curve on F implies that $\mathfrak{A}^*$ has no irrelevant components.*

Proof. If we let $\mathfrak{J}$ denote the ring $\mathfrak{o}^*/\xi_2^*$ then since $\mathfrak{o}^*\xi_2^*$ is prime, $\mathfrak{J}$ is an integral domain, and since $\xi_2^*=0$ is arithmetically normal, $\mathfrak{J}$ *is integrally closed in its quotient field* $\bar{\Sigma}$. This implies that if $\xi_1^* \to \xi_1 \bmod \xi_2^*$, then the principal ideal $\mathfrak{J}\xi_1$ is an intersection of *minimal* ideals in $\mathfrak{J}$, and has no embedded ideals in $\mathfrak{J}$. Since $\mathfrak{A}^*/\xi_2^* = \mathfrak{J}\xi_1$ it follows that $\mathfrak{A}^*$ can have no embedded ideals, q.e.d. We can thus write $\mathfrak{A}^* = [\mathfrak{p}_1^*, \mathfrak{p}_2^*, \cdots, \mathfrak{p}_r^*]$.

We now show that under the conditions we have specified above the construction of Dedekind-Weber can be generalized to the case now under consideration. Let $\omega_1, \omega_2, \cdots, \omega_\nu$ be ν elements of $\mathfrak{o}_0$ which are linearly independent modulo $\mathfrak{A}(=\mathfrak{o}_0(\xi_1, \xi_2))$. If h_i is the smallest integer such that $\omega_i^* = \xi_0^{*h_i}\omega_i$ is an element of $\mathfrak{o}^*$ then ω_i^* is homogeneous of degree h_i. Let us suppose that in $\mathfrak{o}^*$ a congruence of the form

$$P_1(\xi_0^*)\omega_1^* + P_2(\xi_0^*)\omega_2^* + \cdots + P_\nu(\xi_0^*)\omega_\nu^* \equiv 0(\mathfrak{A}^*) \tag{8.1}$$

is valid, where P_i is a polynomial in ξ_0^*. Let $c_1\xi_0^{*k}$ be any term of P_1, so that $c_1\xi_0^{*k}\omega_1^*$ is homogeneous of degree $s=k+h_1$. The homogeneous component of degree s on the left-hand side of (8.1) will then be $c_1\xi_0^{*k}\omega_1^* + c_2\xi_0^{*s-h_2}\omega_2^* + \cdots + c_\nu\xi_0^{*s-h_\nu}\omega_\nu^*$, where $c_i\xi_0^{*s-h_i}$ is a term of degree $s-h_i$ in P_i. Since $\mathfrak{A}^*$ is an H-ideal, we must have $\sum c_i\xi_0^{*s-h_i}\omega_i^* = A\xi_1^* + B\xi_2^*, A, B \subset \mathfrak{o}^*$. If $A = A_m + A_{m-1} + \cdots + A_0$ and $B = B_m + B_{m-1} + \cdots + B_0$, where A_i and B_i are forms of degree i in $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ then under the automorphism $\tau: \xi_0^* \to t\xi_0^*$, $t \in K$ we obtain

$$t^s(A\xi_1^* + B\xi_2^*) = t^{m+1}(A_m\xi_1^* + B_m\xi_2^*) + t^m(A_{m-1}\xi_1^* + B_{m-1}\xi_2^*) + \cdots + t(A_0\xi_1^* + B_0\xi_2^*).$$

Since this is an identity in t, $A\xi_1^* + B\xi_2^* = A_{s-1}\xi_1^* + B_{s-1}\xi_2^*$ and $\sum c_i\xi_0^{*s-h_i}\omega_i^* = A_{s-1}\xi_1^* + B_{s-1}\xi_2^*$. If we divide this last equation by ξ_0^{*s} we find $c_1\omega_1 + c_2\omega_2 + \cdots + c_\nu\omega_\nu \equiv 0\ (\mathfrak{A})$. This implies $c_1 = c_2 = \cdots = c_\nu = 0$. Repeating this process with the other homogeneous components on the left-hand side of (8.1) yields the conclusion that (8.1) implies $P_1 \equiv P_2 \equiv \cdots \equiv P_\nu \equiv 0$.

Conversely, if ω^* is an element of $\mathfrak{o}^*$ which is homogeneous of degree s, then $\omega^*/\xi_0^{*s} = \omega$ is in $\mathfrak{o}_0$ and $\omega - (c_1\omega_1 + c_2\omega_2 + \cdots + c_\nu\omega_\nu) = a\xi_1 + b\xi_2$, a and b in $\mathfrak{o}_0$ and $c_i \in K$, $i=1, 2, \cdots, \nu$. If h is an integer so large that $\xi_0^{*h}\omega$, $\xi_0^{*h}\omega_i^*$, $\xi_0^{*h-1}a$ and $\xi_0^{*h-1}b$ are all in $\mathfrak{o}^*$ then we have

$$\xi_0^{*k-s}\omega^* \equiv \sum_{i=1}^{\nu} c_i\xi_0^{*k-h_i}\omega_i^*(\mathfrak{A}^*). \tag{8.2}$$

We select elements $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ in $\mathfrak{o}^*$ as follows: take $\lambda_1^*=1$, and let λ_2^* be an element in $\mathfrak{o}^*$ of lowest degree of homogeneity, r_2, such that λ_2^* does not satisfy a congruence of the form $\xi_0^{*k-r_2}\lambda_2^* \equiv c_1\xi_0^{*k-r_1}\lambda_1^*$ $(\mathfrak{A}^*)$, where $r_1=0$ is the degree of λ_1^*. We let λ_3^* be a homogeneous element of lowest degree in $\mathfrak{o}^*$, say r_3, such that λ_3^* does not satisfy a congruence of the form $\xi_0^{*k-r_3}\lambda_3^* \equiv c_1\xi_0^{*k-r_1}\lambda_1^* + c_2\xi_0^{*k-r_2}\lambda_2^*$ $(\mathfrak{A}^*)$, $c_i \in K$, for any integer k. We must have $r_2 \leqq r_3$, for any element ω^* in $\mathfrak{o}^*$ of degree $h<r_2$ satisfies a congruence of the form $\xi_0^{*k-h}\omega^* \equiv c_1\xi_0^{*k-r_1}\lambda_1^*$ $(\mathfrak{A}^*)$. In general, if $\lambda_1^*, \lambda_2^*, \cdots, \lambda_i^*$ have been selected, and they are of degrees r_1 $(=0)$, $1 \leqq r_2 \leqq r_3 \leqq \cdots \leqq r_i$ respectively, then we let λ_{i+1}^* be an element of smallest degree of homogeneity, r_{i+1}, which does not satisfy a congruence of the form

$$\xi_0^{*k-r_{i+1}}\lambda_{i+1}^* \equiv c_1\xi_0^{*k-r_1}\lambda_1^* + \cdots + c_i\xi_0^{*k-r_i}\lambda_i^*(\mathfrak{A}^*), \qquad c_i \in K,$$

for any integer k. We of course have $r_{i+1} \geqq r_i$.

This process yields exactly ν functions $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ in $\mathfrak{o}^*$. This statement follows directly from the fact that $\mathfrak{o}_0/\mathfrak{A}$ is of rank ν over K, in view of our considerations on the congruences (8.1) and (8.2). Moreover, if we put $\lambda_i=\lambda_i^*/\xi_0^{*r_i}$, then $\lambda_1, \lambda_2, \cdots, \lambda_\nu$ are linearly independent modulo $\mathfrak{A}$, for if $\lambda_1, \lambda_2, \cdots, \lambda_{i-1}$ are independent mod $\mathfrak{A}$ while $\lambda_i=c_1\lambda_1+\cdots+c_{i-1}\lambda_{i-1}+a\xi_1+b\xi_2$, then for a sufficiently large value of k we have $\xi_0^{*k-r_i}\lambda_i^* \equiv c_1\xi_0^{*k-r_1}\lambda_1^* + \cdots + c_{i-1}\xi_0^{*k-r_{i-1}}\lambda_{i-1}^*(\mathfrak{A}^*)$. The choice of λ_i^* excludes the possibility of such a congruence. In a similar manner we deduce that if ω^* is an element of $\mathfrak{o}^*$ which is homogeneous of degree h, and if, say, r_i is the last element of the sequence $r_1 \leqq r_2 \leqq \cdots \leqq r_\nu$ such that $h \geqq r_i$ then ω^* satisfies a congruence of the form

$$\xi_0^{*k-h}\omega^* \equiv \sum_{j=1}^{i} c_j\xi_0^{*k-r_j}\lambda_j^*(\mathfrak{A}^*) \tag{8.3}$$

where k is an integer. Since $h \geqq r_j$, $j=1, 2, \cdots, i$, we can write

$$\xi_0^{*k-h}\left(\omega^* - \sum_{j=1}^{i} c_j\xi_0^{*h-r_j}\lambda_j^*\right) \equiv 0(\mathfrak{A}^*), \tag{8.4}$$

and since $\mathfrak{A}^*$ has no projectively irrelevant components we have

$$\omega^* - \sum_{j=1}^{i} c_j\xi_0^{*h-r_j}\lambda_j^* = A\xi_1^* + B\xi_2^* \tag{8.5}$$

where A and B are in $\mathfrak{o}^*$.

We can now easily prove the following theorem.

THEOREM 8.2. *The elements $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ form a modular base for $\mathfrak{o}^*$ over*

the ring R^*. In fact, if ω^* is a homogeneous element of degree h in $\mathfrak{o}^*$, and r_i is the last integer r_j such that $h \geqq r_j$, then $\omega^* = P_1\lambda_1^* + \cdots + P_i\lambda_i^*$ where P_j is in R^* and is homogeneous of degree $h - r_j$.

Proof. If $h=0$ the assertion is trivially true. Assume, therefore, that the theorem is true for elements in $\mathfrak{o}^*$ which are homogeneous of degree less than h, and let ω^* be of degree h. We note that the left-hand side of (8.5) is homogeneous of degree h, so that we may assume that the elements A and B on the right-hand side of (8.5) are homogeneous of degree $h-1$. (This follows by the usual application of the mapping $\tau: \xi_0^* \to t\xi_0^*$, $t \in K$.) By induction, then, we have $A = \sum_{j=1}^{i} A_j\lambda_j^*$, $B = \sum_{j=1}^{i} B_j\lambda_j^*$, where A_j and B_j are homogeneous of degree $h - r_j - 1$, $j = 1, \cdots, i$. Hence by (8.5), $\omega^* = \sum P_j\lambda_j^*$ where $P_j = c_j\xi_0^{*h-r_j} + A_j\xi_1^* + B_j\xi_2^*$. Since there are at least ν elements in $\mathfrak{o}^*$ which are linearly independent over R^*, the elements $\lambda_1^*, \lambda_2^*, \cdots, \lambda_\nu^*$ are necessarily independent over R^*, q.e.d.

References

1. G. Castelnuovo, *Alcune proprietà fondamentali dei sistemi lineari di curve tracciate sopra una superficie algebraica*, Memorie di matematica e di fisica della Società Italiana delle Scienze (3) vol. 10 (1896).
2. R. Dedekind and H. Weber, *Theorie der algebraischen Funktionen einer Veränderlichen*, J. Reine Angew. Math. vol. 92 (1882).
3. W. Krull, *Allgemeine Bewertungstheorie*, J. Reine Angew. Math. vol. 167 (1932).
4. H. T. Muhly, *A remark on normal varieties*, Ann. of Math. vol. 42 (1941).
5. W. Schmeidler, *Grundlagen einer Theorie der algebraischen Funktionen mehrerer Veränderlichen*, Math. Zeit. vol. 28 (1928).
6. F. Severi, *Sulla deficienza della serie caratteristica di un sistema lineare di curve appartenenti ad una superficie algebrica*, Atti della Reale Accademia dei Lincei, Rendiconti (5) vol. 12 (1903).
7. B. L. van der Waerden, *On Hilbert's function, series of composition of ideals, and a generalization of a theorem of Bezout*, Proceedings of the Koninklijke Akademie van wetenschappen, Amsterdam vol. 31 (1928).
8. O. Zariski, *Algebraic surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete vol. 3 (1935).
9. ———, *Some results in the arithmetic theory of algebraic varieties*, Amer. J. Math. vol. 61 (1939).
10. ———, *Algebraic varieties over ground fields of characteristic zero*, Amer. J. Math. vol. 62 (1940).
11. ———, *Pencils on an algebraic variety and a new proof of a theorem of Bertini*, Trans. Amer. Math. Soc. vol. 50 (1941).
12. ———, *Normal varieties and birational correspondences*, Bull. Amer. Math. Soc. vol. 48 (1942).

PRINCETON UNIVERSITY,
PRINCETON, N. J.
U. S. NAVAL ACADEMY,
ANNAPOLIS, MD.

GENERALIZATION OF A PROBABILITY LIMIT THEOREM OF CRAMÉR

BY

W. FELLER

1. **Introduction.** Let $\{X_k\}$, $k=1, 2, \cdots$, be mutually independent random variables, that is to say independent real functions in a space in which a probability measure is defined. Let $V_k(x)=\text{Prob}\{X_k \leqq x\}$ be the distribution function of X_k and suppose that the second moment

$$\sigma_k^2 = \int_{-\infty}^{+\infty} x^2 dV_k(x) \tag{1.1}$$

exists; without loss of generality we may then assume that

$$\int_{-\infty}^{+\infty} x dV_k(x) = 0. \tag{1.2}$$

Let

$$S_n = X_1 + \cdots + X_n \tag{1.3}$$

and

$$s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2. \tag{1.4}$$

Finally, let $F_n(x)=\text{Prob}\{S_n \leqq x\}$ be the distribution function of S_n. The independence of the X_n implies that

$$F_1(x) = V_1(x), \qquad F_{n+1}(x) = \int_{-\infty}^{+\infty} F_n(x-y)dV_{n+1}(y). \tag{1.5}$$

The central limit theorem asserts that, under very general conditions,

$$F_n(xs_n) \to \Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-(1/2)y^2}dy, \tag{1.6}$$

uniformly in every finite interval. If, however, with $n\to\infty$ also $x\to\infty$ it follows only that $F_n(xs_n)\to 1$, but it is not even clear whether $1-F(xs_n)$ is of the same order of magnitude as $1-\Phi(x)$. In the Bernoulli case with

$$V_k(x) \equiv V(x) = \begin{cases} 0 & \text{for} \quad x < -p < 0, \\ 1-p & \text{for} \quad -p \leqq x < 1-p, \\ p & \text{for} \quad 1-p \leqq x, \end{cases}$$

Presented to the Society, February 27, 1943; received by the editors March 15, 1943.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

the behavior of $F_n(xs_n)$ when $x=x(n)\to\infty$ has been the object of detailed investigations by A. Khintchine [5](1), N. Smirnoff [7], P. Lévy [6] and M. Fréchet [4]. A more general case has recently been treated by H. Cramér [2]; he considers the case of "equal components" where all $V_k(x)$ are identical, $V_k(x)\equiv V(x)$, and it is moreover supposed that the integral

$$\int_{-\infty}^{+\infty} e^{ax}dV(x) \tag{1.7}$$

exists for all $|a|<a_0\neq 0$. Nothing seems to be known in the case of unequal components. The problem as such has been stated by P. Lévy [6, p. 289].

The purpose of the present paper is to generalize Cramér's result to the case of unequal components. This generalization seems of interest in itself, but it is also to serve as the basis for the establishment of the general form of the so-called law of the iterated logarithm.

We shall consider only the case where $|X_n|$ has an upper bound of the form $o(s_n)$. Since $s_n\to\infty$, this restriction is not very strong. In some cases it is possible to free oneself from this restriction, provided the "tails" $\int_{|x|>\epsilon_n}dV_n(x)$ are sufficiently small. However, in many applications the principle of equivalent sequences of random variables will automatically reduce the consideration to the case of bounded variables. Our proof is based on a straightforward adaptation of Cramér's proof, although we shall avoid the use of characteristic functions. Cramér's proof, in turn, rests on a transformation first used by F. Esscher [3] in connection with an insurance problem. In the special case of equal components our condition $|X_n|=o(s_n)$ is more restrictive than Cramér's condition that the integral (1.7) should exist. On the other hand, our results will be slightly sharper than Cramér's, since it will be possible for us to free ourselves of the logarithmic term in the remainder without introducing a new hypothesis.

2. **Results.** Throughout what follows we shall suppose that the $\{X_k\}$ are subject to the conditions (1.1) and (1.2). Moreover, we shall suppose that (cf. (1.4))

$$|X_k|<\lambda_n s_n, \qquad \text{for } k=1,\cdots,n. \tag{2.1}$$

In most applications the $\{X_k\}$ will form an infinite sequence, and $\lambda_n\to 0$. Condition (2.1) can then be replaced by $|X_n|<\lambda_n s_n$, provided that the sequence λ_n is non-increasing. It should be noted, however, that Theorem 1 is strictly a theorem on a finite sequence of random variables.

In §§3–5 we shall prove

THEOREM 1. *If*

$$0<\lambda_n x<(3-5^{1/2})/4 \tag{2.2}$$

(1) The numbers in brackets refer to the bibliography at the end of the paper.

then

$$1 - F_n(xs_n) = e^{-(1/2)x^2Q_n(x)}\left[\{1 - \Phi(x)\} + \theta\lambda_n e^{-(1/2)x^2}\right]. \tag{2.3}$$

Here θ has an upper bound independent of the sequence $\{X_k\}$ and of n; furthermore

$$Q_n(x) = \sum_{\nu=1}^{\infty} q_{n,\nu}x^\nu \tag{2.4}$$

where the coefficient $q_{n,\nu}$ depends only on the ν first moments of $V_1(x), \cdots, V_n(x)$. If, more particularly,

$$0 < \lambda_n x < 1/12 \tag{2.5}$$

then

$$|\theta| < 9 \tag{2.6}$$

and

$$|q_{n,\nu}| < (1/7)(12\lambda_n)^\nu. \tag{2.7}$$

Moreover, for any $0 < i < j \leqq n$ we have

$$|Q_j(x) - Q_i(x)| \leqq (1/2)(s_j^2 - s_i^2)/s_n^2. \tag{2.8}$$

Before describing the construction of $q_{n,\nu}$ in general we proceed to describe a few implications of Theorem 1. First let us observe that by applying the theorem to the sequence $\{-X_k\}$, we obtain a corresponding theorem for negative x. Since however $Q_n(x)$ depends also on moments of odd order the behavior of $F_n(xs_n)$ for $x \to -\infty$ will, in general, be different.

Since

$$1 - \Phi(x) = (2\pi)^{-1/2}x^{-1}e^{-(1/2)x^2}(1 - \vartheta/x^2) \tag{2.9}$$

with $0 < \vartheta < 1$ we can write (2.3) in the form

$$1 - F_n(xs_n) = (2\pi)^{-1/2}x^{-1}e^{-(1/2)x^2(1+Q_n(x))}\left[1 - \vartheta/x^2 + (2\pi)^{1/2}\theta\lambda_n x\right], \tag{2.10}$$

which shows the order of magnitude of the remainder as compared with that of the main term.

To appreciate the meaning of the theorem consider the representative case of a uniformly bounded sequence $\{X_k\}$ with $s_n^2 \sim n$. Then $\lambda_n = O(n^{-1/2})$. If, then, $x = o(n^{1/6})$ it is seen that $Q_n(x) \to 0$ and $\lambda_n x_n = o(n^{-1/3})$; thus $1 - F_n(xs_n)$ will be of the order of magnitude of $1 - \Phi(x)$. However, if x is of an order of magnitude between $n^{1/6}$ and $n^{1/4}$, the order of magnitude of $1 - F_n(xs_n)$ will be that of $\{1 - \Phi(x)\}e^{-(1/2)q_{n,2}x^2}$; for an x between $n^{1/4}$ and $n^{3/10}$ the term $(1/2)q_{n,3}x^4$ will be significant also, whereas $\sum_{\nu=5}^{\infty} q_{n,\nu}x^{\nu+2} \sim 0$, and so on. Still

$$Q_n(x) = o(1) \qquad \text{for } x = o(n^{1/2}),$$

so that for $x=o(n^{1/2})$ at least $\log(1-F_n(xs_n))\sim\log(1-\phi(x))$. Now if $x\sim n^{1/2}$, the behavior is quite different. Then $(x^2/2)Q_n(x)\sim$const. x^2, and $\log(1-F(xs_n))$ will be asymptotically equivalent to $(\alpha/2)x^2$ with $\alpha\neq 1$; in other words $\log(1-F_n(xs_n))$ will no longer be asymptotically equivalent to $\log(1-\phi(x))$. A similar change in the asymptotic behavior of $1-F_n(xs_n)$ can be observed in the general case and is of fundamental importance in connection with the law of the iterated logarithm.

For use in the same connection we note also the following simple corollary to Theorem 1.

THEOREM 2. *Let $\alpha>0$ be an arbitrary constant and let* $\max(2,\alpha)<x<1/(100\lambda_n)$ and $\lambda_n<1/100$. *Then there exist two constants $0<a_1<a_2<1$, depending only on α, such that*

$$a_1 < \frac{1-F((x+\alpha/x)s_n)}{1-F(xs_n)} < a_2. \tag{2.11}$$

Proof. The theorem is an easy consequence of (2.10). The term in brackets lies obviously between 1/2 and 3/2. From (2.7) we conclude furthermore that

$$\begin{aligned}\left| q_{n,\nu}(x+\alpha/x)^{\nu+2} - q_{n,\nu}x^{\nu+2}\right| &\leqq \left| q_{n,\nu}\right| (\nu+2)(x+\alpha/x)^{\nu+1}\alpha/x \\ &\leqq \frac{3\alpha(\nu+2)}{14}\cdot\left(\frac{24}{100}\right)^{\nu},\end{aligned}$$

so that $\alpha\leqq\left|(x+\alpha/x)^2(1+Q_n(x+\alpha/x))-x^2(1+Q_n(x))\right|\leqq 3\alpha$.

Choosing, say, $a_2=3e^{-\alpha/2}$ and $a_1=(1/5)e^{-3\alpha/2}$ the theorem follows.

Theorem 2 is essentially equivalent to the following more elegant theorem which has been discovered by Khintchine [5] in the case of the binomial and generalized by Cramér [2] to the case of equal components.

THEOREM 3. *If $x\to\infty$, $x=o(1/\lambda_n)$, then*

$$\lim_{n\to\infty}\frac{1-F_n((x+\alpha/x)s_n)}{1-F_n(xs_n)} = \lim_{n\to\infty}\frac{1-\Phi(x+\alpha/x)}{1-\Phi(x)} = e^{-\alpha}.$$

The proof is similar to the preceding one.

We proceed, finally, to describe the *construction of the function $Q_n(x)$. Let*[3]

$$b_{k,\nu} = \int_{-\infty}^{+\infty} x^\nu dV_k(x) \tag{2.12}$$

and define quantities $\gamma_{k,\nu}$ (Thiele's semi-invariants of $V_k(x)$) *by*

$$\log\int_{-\infty}^{+\infty} e^{hy}dV_k(y) = \sum_{\nu=2}^{\infty}\gamma_{k,\nu}\frac{h^\nu}{\nu!}; \tag{2.13}$$

(3) Note that here as in the following the integral actually extends only over the interval $|x|<\lambda_n s_n$.

(we have in particular

$$\gamma_{k,2} = b_{k,2}, \quad \gamma_{k,3} = b_{k,3}, \quad \gamma_{k,4} = b_{k,4} - 3b_{k,2}^2, \tag{2.14}$$
$$\gamma_{k,5} = b_{k,5} - 10b_{k,2}b_{k,3}, \cdots$$

while $\gamma_{k,1}=0$ by (1.2)). *We put*

$$\Gamma_{n,\nu} = \sum_{k=2}^{n} \gamma_{k,\nu}. \tag{2.15}$$

Then

$$\frac{x^2}{2} + \frac{x^2}{2} Q_n(x) = \sum_{\nu=2}^{\infty} \Gamma_{n,\nu} \frac{\nu - 1}{\nu!} h^\nu \tag{2.16}$$

where $h=h(x)$ is defined as the inverse function of

$$x = \frac{1}{s_n} \sum_{\nu=2}^{\infty} \Gamma_{n,\nu} \frac{h^{\nu-1}}{(\nu - 1)!}. \tag{2.17}$$

For the first two coefficients we obtain

$$q_{n,1} = \frac{1}{3s_n^3} \sum_{k=1}^{n} b_{k,3}, \tag{2.18}$$

$$q_{n,2} = \frac{1}{12s_n^4} \sum_{k=1}^{n} b_{k,4} - \frac{1}{4s_n^4} \sum_{k=1}^{n} b_{k,2}^2 - \frac{1}{4s_n^6} \left(\sum_{k=1}^{n} b_{k,3} \right)^2. \tag{2.19}$$

3. **A transformation.** Following Esscher and Cramér we shall introduce a transformation which will reduce $\{V_k(x)\}$ to a new sequence of distribution functions $\{\overline{V}_k(x)\}$ for which the central limit theorem will become applicable. For this purpose we introduce a real parameter $h>0$ which will later on be specialized as a function of x and n.

We put

$$p_k = \int_{-\infty}^{+\infty} e^{hz} dV_k(x), \tag{3.1}$$

and

$$\overline{V}_k(x) = \frac{1}{p_k} \int_{-\infty}^{x} e^{hy} dV_k(y). \tag{3.2}$$

Then $\overline{V}_k(x)$ is again a distribution function. Moreover, $\overline{V}_k(x)$ obviously varies only within a finite interval. Hence the moments

$$\overline{m}_k = \int_{-\infty}^{+\infty} x d\overline{V}_k(x) \tag{3.3}$$

and

$$\bar{\sigma}_k^2 = \int_{-\infty}^{+\infty} (x - \bar{m}_k)^2 d\overline{V}_k(x) = \int_{-\infty}^{+\infty} x^2 d\overline{V}_k(x) - \bar{m}_k^2 \tag{3.4}$$

exist. In a manner analogous to (1.4) and (1.5) we define

$$\overline{M}_n = \bar{m}_1 + \cdots + \bar{m}_n, \tag{3.5}$$

$$\bar{s}_n^2 = \bar{\sigma}_1^2 + \cdots + \bar{\sigma}_n^2, \tag{3.6}$$

and

$$\begin{aligned} \overline{F}_1(x) &= \overline{V}_1(x), \\ \overline{F}_{n+1}(x) &= \int_{-\infty}^{+\infty} \overline{F}_n(x - y) d\overline{V}_{n+1}(y). \end{aligned} \tag{3.7}$$

Now by (3.2)

$$\int_{-\infty}^{x} e^{-hy} d\overline{V}_k(y) = \frac{1}{p_k} \int_{-\infty}^{x} dV_k(y) = \frac{1}{p_k} V_k(x). \tag{3.8}$$

Hence it is seen that

$$F_n(x) = p_1 \cdots p_n \int_{-\infty}^{x} e^{-hy} d\overline{F}_n(y) \tag{3.9}$$

holds for $n=1$; using (1.5) and (3.7) it is seen that (3.9) holds generally. Since $F_n(x)$ is a distribution function we have, letting $x \to \infty$,

$$1 - F_n(x) = p_1 \cdots p_n \int_{x}^{\infty} e^{-hy} d\overline{F}_n(y). \tag{3.10}$$

Now it is our purpose to apply the central limit theorem to $\overline{F}_n(x)$. It will therefore be convenient to write (3.10) in the form

$$1 - F_n(x) = p_1 \cdots p_n \int_{(x-\overline{M}_n)/\bar{s}_n}^{\infty} e^{-h(\overline{M}_n + y\bar{s}_n)} d\overline{F}_n(\overline{M}_n + y\bar{s}_n). \tag{3.11}$$

This formula will be used for $x = \overline{M}_n$ only. Introducing the quantity

$$P_n = \sum_{k=1}^{n} \log p_k \tag{3.12}$$

we deduce from (3.11) finally

$$1 - F_n(\overline{M}_n) = e^{P_n - h\overline{M}_n} \int_{0}^{\infty} e^{-hy\bar{s}_n} d\overline{F}_n(\overline{M}_n + y\bar{s}_n). \tag{3.13}$$

4. **Some estimates.** Operating for the moment in a purely formal manner we have by (2.13) and (3.1)

$$\log p_k = \sum_{\nu=2}^{\infty} \gamma_{k,\nu} \frac{h^\nu}{\nu!}. \tag{4.1}$$

A simple computation shows that (cf. (3.3) and (3.4))

$$\bar{m}_k = \frac{d \log p_k}{dh} = \sum_{\nu=2}^{\infty} \gamma_{k,\nu} \frac{h^{\nu-1}}{(\nu-1)!}, \tag{4.2}$$

$$\bar{\sigma}_k^2 = \frac{d^2 \log p_k}{dh^2} = \sum_{\nu=2}^{\infty} \gamma_{k,\nu} \frac{h^{\nu-2}}{(\nu-2)!}. \tag{4.3}$$

Using (2.16), (3.5), and (3.6) we obtain

$$P_n = \sum_{\nu=2}^{\infty} \Gamma_{n,\nu} \frac{h^\nu}{\nu!}, \tag{4.4}$$

$$\bar{M}_n = \sum_{\nu=2}^{\infty} \Gamma_{n,\nu} \frac{h^{\nu-1}}{(\nu-1)!}, \tag{4.5}$$

$$\bar{s}_n^2 = \sum_{\nu=2}^{\infty} \Gamma_{n,\nu} \frac{h^{\nu-2}}{(\nu-2)!}. \tag{4.6}$$

We now proceed to find majorating power series for these expressions. It follows from (2.1) and (2.12) that for $k=1, \cdots, n$

$$|b_{k,\nu}| \leqq \sigma_k^2 (\lambda_n s_n)^{\nu-2} \leqq (\lambda_n s_n)^\nu. \tag{4.7}$$

Now by (1.2)

$$p_k = 1 + \sum_{\nu=2}^{\infty} b_{k,\nu} \frac{h^\nu}{\nu!}. \tag{4.8}$$

We shall actually use only values $0<\lambda_n s_n h<1/5$, so that p_k remains near 1. It follows from (4.1) and (4.7) that

$$\begin{aligned} \sum_{\nu=2}^{\infty} \gamma_{k,\nu} \frac{h^\nu}{\nu!} &\prec -\log\left\{1 - \frac{\sigma_k^2}{\lambda_n^2 s_n^2} \sum_{\nu=2}^{\infty} \frac{(\lambda_n s_n h)^\nu}{\nu!}\right\} \\ &= -\log\left\{1 - \frac{\sigma_k^2}{\lambda_n^2 s_n^2} (e^{\lambda_n s_n h} - 1 - \lambda_n s_n h)\right\}, \end{aligned} \tag{4.9}$$

where the sign $\prec$ indicates that the power series on the left is majorated by that on the right. Now write for the moment

$$-\log\{1-(e^x-1-x)\}=\sum_{\nu=2}^{\infty}\rho_\nu\frac{x^\nu}{\nu!} \tag{4.10}$$

so that $\rho_2=1$, $\rho_3=1$, $\rho_4=4$, $\rho_5=11$, $\rho_\nu\geqq 0$. Differentiating (4.10) twice we obtain

$$\frac{xe^x+1}{(2+x-e^x)^2}=\sum_{\nu=2}^{\infty}\rho_\nu\frac{x^{\nu-2}}{(\nu-2)!}. \tag{4.11}$$

Substituting here $x=1/2$ we obtain easily that

$$\sum_{\nu=6}^{\infty}\rho_\nu\frac{1}{(\nu-2)!}(1/2)^{\nu-2}<3/10,$$

whence

$$0\leqq\rho_\nu/(\nu-2)!\leqq(1/2)\cdot 2^{\nu-2}\qquad\text{for }\nu\geqq 3. \tag{4.12}$$

Comparing (4.9) with (4.10) and noting that $\sigma_k^2<\lambda_n^2 s_n^2$ we deduce readily that

$$|\gamma_{k,\nu}|\leqq(\sigma_k^2/(\lambda_n s_n)^2)\rho_\nu(\lambda_n s_n)^\nu,\qquad \nu\geqq 3, \tag{4.13}$$

or by (4.12)

$$(1/(\nu-2)!)\,|\gamma_{k,\nu}|\leqq(1/2)\sigma_k^2(2\lambda_n s_n)^{\nu-2},\qquad \nu\geqq 3. \tag{4.14}$$

Accordingly we have by (2.15)

$$(1/(\nu-2)!)\,|\Gamma_{n,\nu}|\leqq(s_n^2/2)(2\lambda_n s_n)^{\nu-2},\qquad \nu\geqq 3, \tag{4.15}$$

whilst $\Gamma_{n,2}=s_n^2$; this allows us to deduce majorating expressions for P_n, $\overline{M}_n$ and $\bar{s}_n^2$ (cf. (4.4)–(4.6)).

5. **Application of the central limit theorem.** Using (4.14) and (4.2) it follows that

$$|\bar{m}_k|\leqq\sigma_k^2 h\left(1+\frac{1}{2}\frac{\lambda_n s_n h}{1-2\lambda_n s_n h}\right)\leqq\lambda_n^2 s_n^2 h\left(1+\frac{1}{2}\frac{\lambda_n s_n h}{1-2\lambda_n s_n h}\right); \tag{5.1}$$

observing that $\Gamma_{n,2}=s_n^2$ it follows similarly from (4.6) that

$$|\bar{s}_n^2-s_n^2|\leqq\sum_{\nu=3}^{\infty}|\Gamma_{n,\nu}|\frac{h^{\nu-2}}{(\nu-2)!}\leqq s_n^2\cdot\frac{\lambda_n s_n h}{1-2\lambda_n s_n h}. \tag{5.2}$$

Now, by definition, $\overline{V}_k(x)$ $(k=1,\cdots,n)$ varies only within the interval $|x|<\lambda_n s_n$. Accordingly, if we restrict $\lambda_n s_n h$ to an interval $0<\lambda_n s_n h<c<1/2$, $\overline{V}_k(x)$ will not vary outside of an interval centered at $\bar{m}_k$ and of length $O(\lambda_n\bar{s}_n)$. It follows (A. C. Berry [1, Theorem 4]) that (cf. (1.6))

$$|\overline{F}_n(\overline{M}_n+x\bar{s}_n)-\Phi(x)|=O(\lambda_n).$$

More precisely let us suppose that, say,

(5.3) $$0 < \lambda_n s_n h < 1/10.$$

Then

$$|\bar{s}_n^2/s_n^2 - 1| < 1/8, \quad \text{and} \quad |\bar{m}_k| < (17/160)\lambda_n s_n.$$

Therefore for $k=1, \cdots, n$

$$\int_{|x|>(6/5)\lambda_n\bar{s}_n} d\bar{V}_k(x+\bar{m}_k) = \frac{1}{p_k}\int_{|x-\bar{m}_k|>(6/5)\lambda_n\bar{s}_n} e^{hx}d\bar{V}_k(x)$$

$$\leqq \frac{1}{p_k}\int_{|x|>\lambda_n s_n} e^{hx}dV_k(x) = 0.$$

Accordingly, using Berry's theorem, we have

(5.4) $$\bar{F}_n(\bar{M}_n + x\bar{s}_n) = \Phi(x) + R_n(x)$$

with

(5.5) $$|R_n(x)| < (13/3)\lambda_n.$$

Introducing (5.4) into (3.13) we obtain

(5.6) $$1 - F_n(\bar{M}_n) = e^{P_n - h\bar{M}_n}\left\{(2\pi)^{-1/2}\int_0^\infty e^{-h\bar{s}_n y-(1/2)y^2}dy + \int_0^\infty e^{-h\bar{s}_n}dR_n(y)\right\}.$$

Now, using (5.5) we obtain

$$\left|\int_0^\infty e^{-h\bar{s}_n y}dR_n(y)\right| = \left| - R_n(0) + h\bar{s}_n\int_0^\infty e^{-h\bar{s}_n y}R_n(y)dy\right| \leqq (26/3)\lambda_n$$

(which result can be derived also by means of the second mean value theorem). The last inequality together with (5.6) shows that

(5.7) $$1 - F_n(\bar{M}_n) = e^{P_n - h\bar{M}_n}\left\{e^{(1/2)(h\bar{s}_n)^2}(1 - \Phi(h\bar{s}_n)) + \theta_1\lambda_n\right\},$$

where

(5.8) $$|\theta_1| < 26/3.$$

We now have to determine h so that

(5.9) $$\bar{M}_n \equiv \sum_{\nu=2}^{\infty}\Gamma_{n,\nu}\frac{h^{\nu-1}}{(\nu-1)!} = xs_n.$$

We write the inverse series formally,

(5.10) $$h = \frac{1}{s_n}\sum_{\nu=1}^{\infty}c_{n,\nu}x^\nu = \frac{1}{s_n}\left(x - \frac{x^2}{2s_n^2}\sum_{k=1}^{n}b_{k,3} + \cdots\right).$$

Using Cauchy's principle for majorating series and the estimate (4.15) it is

readily seen that the formal inversion of (5.9) is possible at least for all values of x for which the inversion of

$$x = s_n h - \frac{1}{4}\sum_{\nu=3}^{\infty}(2\lambda_n)^{\nu-2}(s_n h)^{\nu-1} \tag{5.11}$$

is possible. The substitution

$$\lambda_\nu x = u, \qquad \lambda_n s_n h = z \tag{5.12}$$

reduces (5.11) to

$$u = z - \frac{1}{8}\sum_{\nu=2}^{\infty}(2z)^\nu = z - \frac{1}{2}\,\frac{z^2}{1-z}\,; \tag{5.13}$$

the inverse function

$$z = \frac{1 + 2u - |\,1 - 6u + 4u^2\,|^{1/2}}{5} = u + \sum_{\nu=2}^{\infty} c_\nu^* u^\nu \tag{5.14}$$

is regular in the interval (2.2). Using an inductive argument we readily deduce from the series-expansion in (5.13) that

$$\left.\frac{d^\nu z}{du^\nu}\right|_{u=0} \geqq 0$$

for all ν. Hence $c_\nu^* \geqq 0$; we have in particular $c_2^* = 1/2$ and $c_3^* = 3/2$. For $u = 1/6$ we obtain $z = 1/5$ and hence

$$\sum_{\nu=1}^{\infty} c_\nu^* 6^{-\nu} = 1/80;$$

accordingly we have for $\nu \geqq 2$

$$0 \leqq c_\nu^* \leqq (1/12)6^{\nu-1}.$$

Returning to the original variables x and h it follows that for $\nu \geqq 2$

$$|\,c_{n,\nu}\,| \leqq (1/12)(6\lambda_n)^{\nu-1} \tag{5.15}$$

and

$$s_n h - x < \frac{x}{12}\,\frac{6\lambda_n x}{1 - 6\lambda_n x}\,. \tag{5.16}$$

We note in particular that the restriction of x to the interval (2.5) implies that

$$\lambda_n s_n h \leqq 13/144 < 1/10 \tag{5.17}$$

so that (5.3) holds; this justifies the numerical estimate of §4.

For $u = (3 - 5^{1/2})/4$ we obtain from (5.14) $z = (1/2)(1 - 5^{-1/2}) < 1/2$. Hence

to values of x in the interval (2.2) there correspond values of $h<1/(2\lambda_n s_n)$; by (4.15) these values are within the interval of convergence of the power series

$$h\overline{M}_n - P_n = \sum_{\nu=2}^{\infty} \Gamma_{n,\nu} \frac{\nu-1}{\nu!} h^\nu. \tag{5.18}$$

Hence we can write

$$h\overline{M}_n - P_n = \frac{x^2}{2} + \frac{x^2}{2} Q_n(x) = \frac{x^2}{2} + \frac{x^2}{2} \sum_{\nu=1}^{\infty} q_{n,\nu} x^\nu, \tag{5.19}$$

where the last series converges at least within (2.2). Now define a new power series

$$Q_n^*(x) = \sum_{\nu=2}^{\infty} q_{n,\nu}^* x^\nu \tag{5.20}$$

by substituting

$$s_n^* h^* = x + \frac{x}{12} \frac{6\lambda_n x}{1 - 6\lambda_n x} \tag{5.21}$$

into

$$\frac{x^2}{2} + \frac{x^2}{2} Q_n^*(x) = \frac{(s_n^* h^*)^2}{2} + \frac{s_n^{*2} h^{*2}}{6} \sum_{\nu=3}^{\infty} (2\lambda_n s_n^* h^*)^{\nu-2}. \tag{5.22}$$

It has been shown that the series (5.21) and (5.22) are dominating series for $s_n h$ and for $h\overline{M}_n - P_n$, respectively. Hence (5.22) dominates the series (5.19) so that $|q_{n,\nu}| \leqq q_{n,\nu}^*$. However, an easy computation shows that $1+Q_n^*(1/12\lambda_n)<1.26$ and that $q_{n,1}^* = 5\lambda_n/3$; it follows that $q_{n,\nu}^* < (1/7)(12\lambda_n)^\nu$, which proves that $Q_n(x)$ has the properties stated in Theorem 1.

Now by (5.7), (5.9) and (5.19)

$$1 - F_n(xs_n) = e^{(1/2)x^2(1+Q_n(x))} \left\{ e^{(1/2)(h\bar{s}_n)^2} (1 - \Phi(h\bar{s}_n)) + \theta_1 \lambda_n \right\}. \tag{5.23}$$

It remains to show that here $h\bar{s}_n$ can be replaced by $x = \overline{M}_n/s_n$. Comparing the power series for $h\bar{s}_n$ and $\overline{M}_n/s_n$ it is seen that the terms of first and second order in h are the same; hence $h\bar{s}_n = x + O(x^3\lambda_n^2)$; more precisely we have

$$|h\bar{s}_n - x| < 2.45\lambda_n^2 x^3, \tag{5.24}$$

provided only that (2.5) holds. Now let for the moment

$$f(t) = e^{(1/2)t^2} \{1 - \Phi(t)\}. \tag{5.25}$$

Then

$$f'(t) = -(2\pi)^{-1/2} + te^{(1/2)t^2}\{1 - \Phi(t)\},$$

and a well known asymptotic expansion shows that

$$0 \leqq -f'(t) \leqq 1/(2\pi)^{1/2}t^2.$$

Accordingly by (5.24)

$$|f(hs_n) - f(x)| < \frac{2.45}{(2\pi)^{1/2}} \frac{\lambda_n^2 x^2}{(x - 2.6\lambda_n^2 x^2)^2} \leqq \lambda_n^2 x.$$

Thus we have by (5.23) and (5.25)

$$1 - F_n(xs_n) = e^{-(1/2)x^2(1+Q_n(x))}\{f(x) + \theta_1\lambda_n + \vartheta\lambda_n\},$$

where $|\vartheta| < \lambda_n x$ and $|\theta_1| < 26/3$ (cf. (5.8)).

This proves the main part of Theorem 1. It remains only to prove (2.8). Now it follows from (2.15) and (2.16) that

$$x^2(Q_j(x) - Q_i(x)) = \sum_{\nu=2}^{\infty} \frac{h^\nu}{(\nu-2)!\nu} \sum_{k=i+1}^{j} \gamma_{k,\nu}.$$

Accordingly, using (4.14),

$$x^2|Q_j(x) - Q_i(x)| \leqq \frac{h^2}{4} \sum_{\nu=2}^{\infty} 2(\lambda_n s_n h)^{\nu-2} \sum_{k=i+1}^{j} \sigma_k^2$$

$$= \frac{h^2}{4}(s_j^2 - s_i^2) \frac{1}{1 - 2\lambda_n s_n h}.$$

This implies (2.8) since $s_n h/x < 13/12$ and $\lambda_n s_n h < 1/10$ by (5.16) and (5.17), respectively.

Bibliography

1. A. C. Berry, *The accuracy of the Gaussian approximation to the sum of independent variates*. Trans. Amer. Math. Soc. vol. 49 (1941) pp. 122–136.

2. H. Cramér, *Sur un nouveau théoreme-limite de la théorie des probabilités*. Actualités Scientifiques, No. 736, Lecture at the Colloque consacré à la Théorie des Probabilités. III. Paris, 1938, pp. 5–23.

3. F. Esscher, *On the probability function in the collective theory of risk*. Skandinavisk Aktuarietidskrift vol. 15 (1932) pp. 175–195.

4. M. Fréchet, *Recherches théoriques modernes sur la théorie des probabilités*. Livre I, Paris, 1937, chap. 2, §5.

5. A. Khintchine, *Ueber einen neuen Grenswertsatz der Wahrscheinlichkeitsrechnung*. Math. Ann. vol. 101 (1929) pp. 745–752.

6. P. Lévy, *Théorie de l'addition des variables aléatoires*. Paris, 1937, chap. 8.

7. N. Smirnoff, *Ueber Wahrscheinlichkeiten grosser Abweichungen*. Recueil Société Mathématique Moscou (Matematičeski Zbornik) vol. 40 (1933) pp. 441–454.

Brown University,
Providence, R. I.

THE GENERAL FORM OF THE SO-CALLED LAW OF THE ITERATED LOGARITHM

BY
W. FELLER

1. **Introduction.** The purpose of this paper is to sharpen Kolmogoroff's celebrated law of the iterated logarithm in various directions and to give best results. For the convenience of the reader, we shall link the statement of our problems with an account of the known results.

1.1. It seems that the law of the iterated logarithm traces its origin to a special problem in number theory. From the viewpoint of the theory of probability, as we shall see, this problem concerns only a very special case in which the general features are hardly visible. Nevertheless, this is the only case in which so far a final result has been achieved.

Let p be a point of the interval (0, 1) and let $p=p_1p_2p_3\cdots$ be its binary expansion. (We shall be concerned with statements of the "almost everywhere" type, so ambiguity will do no harm.) Let

$$X_k(p) = \begin{cases} +1 & \text{if } p_k = 1, \\ -1 & \text{if } p_k = 0. \end{cases}$$

Then $S_n=X_1+\cdots+X_n$ is the excess frequency of occurrence of the digit 1 among the first n places in the expansion of p. The strong law of large numbers (Borel, Cantelli) asserts that almost everywhere $S_n=o(n)$. The following enumeration of sharper results indicates the historic development of the problem. It is, of course, understood that all assertions hold true only for almost all points p.

(i) Hausdorff, [6], 1913:

$$S_n = o(n^{1/2+\epsilon}), \qquad \epsilon > 0.$$

(ii) Hardy-Littlewood, [4], 1914:

$$S_n = O((n \log n)^{1/2}).$$

(iii) Steinhaus, [16], 1922:

$$\limsup_{n\to\infty} S_n/(2n \log n)^{1/2} \leqq 1.$$

(iv) Khintchine, [7], 1923:

$$S_n = O((n \log \log n)^{1/2}).$$

Presented to the Society, February 27, 1943; received by the editors March 15, 1943.

(v) Khintchine, [8], 1924:

$$\limsup_{n\to\infty} S_n/(2n \log\log n)^{1/2} = 1.$$

In formulating newer results we shall, following P. Lévy, say that a function $\phi(t)$ belongs to the upper class ($\in \mathcal{U}$) if, for almost all p, there exist only finitely many n such that $S_n > n^{1/2}\phi(n)$; and we shall say that $\phi(t)$ belongs to the lower class[1] ($\in \mathcal{L}$) if, for almost all p, the inequality $S_n > n^{1/2}\phi(n)$ is satisfied for infinitely many n. Khintchine's result (v) then reads:

$$a(2 \log\log t)^{1/2} \begin{matrix} \in \mathcal{U} & \text{if} \quad a > 1, \\ \in \mathcal{L} & \text{if} \quad a < 1. \end{matrix}$$

(vi) P. Lévy, [11], 1933:

$$(2 \log\log t + a \log\log\log t)^{1/2} \begin{matrix} \in \mathcal{U} & \text{if} \quad a > 3, \\ \in \mathcal{L} & \text{if} \quad a \leqq 1 \end{matrix}$$

(with a gap for $1 < a \leqq 3$).

(vii) Kolmogoroff-Erdös[2], (1937–1942): If $\phi(t)$ is non-decreasing, then[3]

$$\phi(t) \begin{matrix} \in \mathcal{U} \\ \in \mathcal{L} \end{matrix} \quad \text{if} \quad \int^{\infty} \frac{1}{t}\phi(t)e^{-(1/2)\phi^2(t)}dt \begin{matrix} \in \mathcal{C} \\ \in \mathcal{D}. \end{matrix}$$

The last result contains, of course, the preceding ones. It gives, in a most elegant way, a complete solution for the binomial case with probabilities 1/2. It is our purpose to find a similar solution for arbitrary independent random variables, that is to say, for arbitrary bounded real functions X_k in arbitrary spaces. As a matter of fact, the above solution is so specialized that it fails already in the next simplest problem concerning the frequency of occurrence of digits in the ordinary decimal expansion.

1.2. For a general formulation of the problem consider an infinite sequence of spaces $\{E_k\}$, $k=1, 2, \cdots$, in which probability-measures are defined[4]. Let $E = E_1 \times E_2 \times \cdots \times E_k \times \cdots$ be their infinite combinatorial product

---

[1] According to the law of 0 or 1 each $\phi(t)$ belongs to one of these classes. We shall not use this result in the sequel.

[2] Kolmogoroff's result has been stated, without proof, in P. Lévy's book [12]. Erdös [2] proved the result completely. Before, J. Ville [17] had proved that convergence of the integral is a sufficient condition for the upper class. It is the necessity of this condition which presents the real difficulties. Most of the results of the present paper have been obtained before publication of Erdös' paper.

[3] Here and in the following $\mathcal{C}$ and $\mathcal{D}$ stand for "converges" and "diverges," respectively.

[4] In each E_k there is a Borel family of sets $\mathfrak{F}_k$ such that $E_k \in \mathfrak{F}_k$; on $\mathfrak{F}_k$ we have a completely additive set-function $\Pr(\Gamma)$ with $0 \leqq \Pr(\Gamma) \leqq 1$ and $\Pr(E_k) = 1$.

and let a probability-measure in E be defined in the customary manner(5). Let $X_k = X_k(p_k)$ be a real-valued function of a point p_k varying in E_k; then(6)

$$S_n = X_1 + \cdots + X_n \tag{1.1}$$

is a function of the point $p = p_1 \times p_2 \times \cdots \in E$. Let

$$V_k(x) = \Pr\{X_k \leqq x\}, \qquad -\infty < x < \infty, \tag{1.2}$$

be the distribution function of X_k. Then the distribution function $F_n(x)$ of S_n is found from

$$F_{n+1}(x) = \int_{-\infty}^{+\infty} F_n(x - y)dV_{n+1}(y).$$

We shall suppose that

$$\sigma_k^2 = \int_{-\infty}^{+\infty} x^2 dV_k(x) \tag{1.3}$$

exists; replacing, if necessary, X_k by $X_k + c_k$ we can without loss of generality assume that

$$\int_{-\infty}^{+\infty} x dV_k(x) = 0, \qquad k = 1, 2, \cdots. \tag{1.4}$$

The second moment of $F_n(x)$ is

---

(5) We shall actually consider only sets of the type $\Gamma = \cdots \times \Gamma_{i_1} \times \cdots \times \Gamma_{i_2} \times \cdots \times \Gamma_{i_n} \times \cdots$, where $\Gamma_{i_k} \in \mathfrak{F}_{i_k}$ and the dots stand for E_n. Of course, $\Pr(\Gamma) = \Pr(\Gamma_{i_1}) \cdots \Pr(\Gamma_{i_n})$. If $\Gamma = (\Gamma_{i_1}, \cdots, \Gamma_{i_n})$ and $\Delta = (\Delta_{j_1}, \cdots, \Delta_{j_m})$ are two such sets, they will be called *independent*, if $i_\sigma \neq j_\sigma$; then $\Pr(\Gamma\Delta) = \Pr(\Gamma)\Pr(\Delta)$. For our purposes it is sufficient to have a measure-theory for finite combinatorial products and the corresponding notion of sets of measure zero in the infinite product space.

(6) In this formulation it becomes clear that, as far as *mathematics* is concerned, we are dealing with a problem in real function theory in product spaces. The sense, and importance, of our investigation for probability and statistics is, it is hoped, obvious. The subscript k stands, of course, for the order number of the experiment or observation. E_k, the kth label space, is the abstract expression for the set of all thinkable results of the kth experiment, each result being represented by a point $p_k \in E_k$. (In practice these results will be: head-or-tail; the position of the roulette; the position of a particle subjected to diffusion or, more generally, the phase-space in physics; or the complex composition with respect to age, claims, risk, and so on, of the stock of policies of an insurance company.) X_k will be the gain or any other characteristic in the kth experiment and we are interested in the fluctuations of the accumulated value of that characteristic after n experiments. The epistemological problem of what "probability" means has no more to do with theory or practice than the corresponding problem of space conception in geometry.

It may be remarked that we do not attach any importance to the possibility that the E_k are not similar to each other. The fact of the situation is that the structure of E_k simply nowhere appears.

$$s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2. \tag{1.5}$$

For obvious reasons only the case where

$$s_n \to \infty \tag{1.6}$$

presents a problem.

We shall, following P. Lévy, say that *a sequence* $\{\phi_n\}$ *belongs to the lower class* ($\in \mathcal{L}$) (*with respect to* $\{X_k\}$), *if, for almost all* $p \in E$, *there exist infinitely many* n *such that*

$$S_n > s_n \phi_n; \tag{1.7}$$

and to the upper class ($\in \mathcal{U}$) *if, for almost all* $p \in E$, *there exist only finitely many* n *such that* (1.7) *holds.* (It will be noticed that Theorem 8 actually contains a sharpening of the notion of upper class: in defining the upper class we could as well replace (1.7) by (2.19).)

1.3. The law of the iterated logarithm states: *if each* X_k *is bounded and*

$$\text{l.u.b.} \mid X_k \mid = o(s_n/\log_2^{1/2} s_n) \tag{1.8}$$

uniformly in $k = 1, 2, \cdots, n$, *then*

$$(a \log_2 s_n)^{1/2} \begin{cases} \in \mathcal{U} & \text{if } a > 2, \\ \in \mathcal{L} & \text{if } a < 2. \end{cases} \tag{1.9}$$

This is the most general result known; it was proved by A. Kolmogoroff [10], after important special cases had been solved by A. Khintchine [9]. A new proof of (1.9) under more restrictive conditions than (1.8) has been given by P. Lévy [12]. Under much stronger conditions Cantelli [1] has proved a result which is sharper than (1.9): suppose that for some $c > 2$

$$\int_{-\infty}^{+\infty} \mid x \mid^c dV_k(x), \qquad k = 1, 2, \cdots,$$

exists, and that uniformly in k

$$\int_{-\infty}^{+\infty} x^2 dV_k(x) \geqq a_0, \qquad \int_{-\infty}^{+\infty} \mid x \mid^c dV_k(x) \leqq a_1;$$

then (even not assuming (1.8))

$$(2 \log_2 s_n + a \log_3 s_n)^{1/2} \begin{cases} \in \mathcal{U} & \text{if } c > 3, \\ \in \mathcal{L} & \text{if } c \leqq 1; \end{cases}$$

the gap between 1 and 3 is left open. (Cf. above P. Lévy's result vi.)

There is a most surprising result due to Marcinkiewicz and Zygmund [15] which is important for the proper understanding of our problem. They showed

that if (1.8) be replaced by the only slightly less restrictive condition:

$$\text{l.u.b.} \mid X_k \mid < \epsilon(s_n/\log_2^{1/2} s_n) \tag{1.10}$$

(ϵ arbitrary but fixed), then (1.9) becomes *false*. In fact, they constructed a sequence $\{X_k\}$ satisfying (1.10) but for which

$$\limsup_{n\to\infty} (S_n/s_n(2 \log_2 s_n)^{1/2}) < 1. \tag{1.11}$$

1.4. In the present paper we shall give necessary and sufficient conditions for a sequence $\{\phi_n\}$ to belong to the upper or lower class. It will be seen that our criterion assumes the simple form of Kolmogoroff's criterion (vii) if the $\{X_k\}$ are subject to the condition(7) $|X_k| = O(s_n/\phi_n^2)$ (Theorems 2 and 6). As this condition is gradually relaxed the criterion assumes an increasingly complicated form. Thus, if $|X_k| = O(s_n/\phi_n^3)$ the criterion will involve third moments of the distribution functions $V_k(x)$; this implies a remarkable lack of symmetry, since now $\{\phi_n\}$ may belong to the upper class for the sequence $\{X_k\}$ and to the lower class for $\{-X_k\}$ (Theorem 3). If it is assumed only that $|X_k| = O(s_n/\phi_n^{5/3})$ the fourth moments of $\{V_k(x)\}$ will be of influence also, and so on (Theorem 5). Finally, when the upper bounds of $|X_k|$ are of the order $C(s_n/\phi_n)$ all moments of $\{V_k(x)\}$ enter into consideration; when C is very small, the dominant influence will still be exercised by the second moments; however, as C increases, the dominant role gradually shifts to moments of an increasingly higher order. It is thus seen that if the upper bounds of $|X_k|$ increase faster than s_n/ϕ_n, we shall be confronted with an altogether new situation and any criterion would necessarily be of a quite different form(8). This domain is not investigated in the present paper. The criterion given in Theorem 1 still covers the case (1.10) (with $\epsilon = 1/200$) and thus explains the phenomenon detected by Marcinkiewicz and Zygmund by giving the precise limit between the upper and the lower class.

It will be seen that the present paper considers only individually bounded variables X_k. It is hardly necessary to point out that most sequences which occur in standard applications can be reduced to this case by the customary method of equivalent sequences (truncating the X_k at values M_k such that $\sum \Pr\{|X_k| > M_k\}$ converges). This method makes it evident, for example, that Kolmogoroff's law of the iterated logarithm is applicable if

---

(7) For a proper understanding of the theorems (as well as for a comparison with Kolmogoroff's condition (1.8)) it should be noticed that we shall be interested usually in sequences $\{\phi_n\}$ near the borderline between the upper and the lower class. That is to say, roughly speaking ϕ_n will increase like $(2 \log \log s_n)^{1/2}$.

(8) This observation is also borne out by results of P. Lévy [13] and Marcinkiewicz [15], who investigated the analogue to the law of the iterated logarithm in the special case where $cx^{-\alpha} \leq 1 - V_k(x) \leq Cx^{-\alpha}$, with $0 < \alpha < 1$ and two constants $C > c > 0$.

$V_k(x) \equiv V(x)$ where $V(x)$ has a finite moment of order $2+\epsilon > 2$(9). Actually it is possible to extend also the exact criteria of the present paper to very large classes of sequences $\{X_k\}$ of unbounded variables. An example of such a class (containing most usual sequences) is given in the appendix. In the construction it seemed desirable to aim at greater simplicity rather than at greatest possible generality.

Our proofs are direct and do not presuppose previous work on the subject. Beyond standard tools of measure theory we shall use only a theorem describing the behavior of $F_n(x)$ for large x; this theorem is a sharpening of the classical central limit theorem of probability and has been proved in a previous paper(10).

2. **Theorems.** *Throughout the sequel X_k will stand for a real function in E_k* (see §1.2); *S_n is defined by* (1.1) *and $V_k(x)$ by* (1.2); *it shall be assumed that* (1.3), (1.4), *and* (1.6) *hold. Moreover we shall suppose that there exists a sequence $\lambda_n \downarrow 0$ such that*(11)

$$\text{l.u.b.} \mid X_n \mid < \lambda_n s_n, \tag{2.1}$$

where s_n is defined by (1.5). *Of the sequence $\{\phi_n\}$ we shall assume that*(12)

$$2 < \phi_1 \leqq \phi_2 \leqq \phi_3 \leqq \cdots . \tag{2.2}$$

2.1. For the most general case of our criterion we shall require a function $Q_n(x)$ which regulates the behavior of $F_n(xs_n)$ for large x. Define, in a formal way, quantities $\gamma_{k,\nu}$ and $\Gamma_{n,\nu}$ by

$$\log \int_{-\infty}^{+\infty} e^{\lambda y} dV_k(y) = \sum_{\nu=2}^{\infty} \gamma_{k,\nu} \frac{\lambda^\nu}{\nu!}$$

and

---

(9) This seems to have escaped the attention of Hartman and Wintner [5], in their strong remarks about the inapplicability of Kolmogoroff's law to sequences occurring in standard applications. Actually it suffices to suppose that $\int_{-\infty}^{+\infty} x^2 |\log|x|\,|^{1+\epsilon} dV(x) = A$ is finite for some $\epsilon > 0$. Then $\sum_{k=100}^{\infty} \int_{|x|>k^{1/2}/\log\log k} dV(x) < A \sum_{k=100}^{\infty} (\log\log s_k)^4/k \log^{1+\epsilon} k$ converges; the truncated variables obey the law of the iterated logarithm and the variances of their sums are asymptotically equivalent to s^2. Hartman and Wintner [5] have devised a slightly more efficient method of truncation which shows that it is sufficient to suppose that $\int_{-\infty}^{+\infty} x^2 dV(x)$ exists. It does *not* follow that in this case also our criteria hold.

(10) Feller [3], in the sequel referred to as L.T. The present paper should be readable without knowledge of L.T.

(11) The monotony of $\{\lambda_n\}$ is used exclusively to assure that l.u.b. $|X_k| < \lambda_n s_n$ for $k = 1, 2, \cdots, n$.

(12) It seems only natural to assume monotony of $\{\phi_n\}$. The restriction is unessential and the proofs become even slightly simpler if, instead, it is assumed that the variation of $\{\phi_n\}$ in any interval is small as compared with $1/\phi_n$. However, abrupt changes of ϕ_n are interesting and considered in the sequel. If $\{\phi_n\}$ remains bounded, the theorems become trivial. The assumption $\phi_n > 2$ is convenient and presents no loss of generality.

$$\Gamma_{n,\nu} = \sum_{k=1}^{n} \gamma_{k,\nu}$$

($\gamma_{k,1}=0$ by (1.4); only small values of h will be considered and it is understood that the $\gamma_{k,\nu}$ are real). Then

$$\frac{x^2}{2}\{1+Q_n(x)\} \equiv \sum_{\nu=2}^{\infty} \Gamma_{n,\nu}\frac{\nu-1}{\nu!}h^\nu$$

where $h=h(x)$ is the inverse function of

$$x = \frac{1}{s_n}\sum_{\nu=2}^{\infty}\Gamma_{n,\nu}\frac{h^{\nu-1}}{(\nu-1)!}.$$

It has been shown in L.T. that, if $x\lambda_n<1/12$, $Q_n(x)$ is an analytic function

$$Q_n(x) = \sum_{\nu=1}^{\infty} q_{n,\nu}x^\nu \tag{2.3}$$

and

$$|q_{n,\nu}| < (1/7)(12\lambda_n)^\nu. \tag{2.4}$$

In general, $q_{n,\nu}$ depends on the first $\nu+2$ moments of $V_1(x), \cdots, V_n(x)$,

$$b_{k,\rho} = \int_{-\infty}^{+\infty} x^\rho dV_k(x), \qquad \rho = 1, \cdots, \nu+2. \tag{2.5}$$

It follows from the main theorem of L.T. that, under the conditions of the present paper, the ratio of $F(s_n\phi_n)$ to $\phi_n^{-1}\{\exp -(1/2)\phi_n^2[1+Q_n(\phi_n)]\}$ remains between two positive constants. This explains the importance of the function $Q(x)$ for the present investigation.

2.2. In the following theorems the main conditions ((2.6), (2.8), (2.12), and so on) depend on the sequence $\{\phi_n\}$ which is to be tested. Although these conditions seem most natural, it is sometimes more advantageous to replace them by conditions which do not explicitly depend on $\{\phi_n\}$. Such alternative conditions will be formulated in Theorem 11.

THEOREM 1. *If*

$$\lambda_n \leqq 1/200\phi_n \tag{2.6}$$

then $\{\phi_n\}\in\mathcal{U}(\mathcal{L})$ *if, and only if*[12],

$$\sum_n \frac{\sigma_n^2}{s_n^2}\phi_n \exp\{-(1/2)\phi_n^2[1+Q_n(\phi_n)]\} \in \mathcal{C}(\mathcal{D}). \tag{2.7}$$

[12] Cf. footnote 3.

Postponing the proof of the theorem we mention a few corollaries.

THEOREM 2. *If*

$$\lambda_n = O(1/\phi_n^3) \tag{2.8}$$

then $\{\phi_n\} \in \mathcal{U}(\mathcal{L})$ *if, and only if,*

$$\sum_n \frac{\sigma_n^2}{s_n^2} \phi_n e^{-(1/2)\phi_n^2} \in \mathcal{C}(\mathcal{D}). \tag{2.9}$$

In fact, using (2.4) it is seen that (2.8) implies $\phi_n^2 Q_n(\phi_n) = O(1)$ and Theorem 2 follows from Theorem 1. Similarly one proves another corollary:

THEOREM 3. *If*

$$\lambda_n = O(1/\phi_n^2) \tag{2.10}$$

then $\{\phi_n\} \in \mathcal{U}(\mathcal{L})$ *if, and only if,*

$$\sum_n \frac{\sigma_n^2}{s_n^2} \phi_n \exp\left\{-(1/2)\phi_n^2 - (\phi_n^3/6s_n^3)\sum_{k=1}^n b_{k,3}\right\} \in \mathcal{C}(\mathcal{D}) \tag{2.11}$$

where $b_{k,3}$ *is defined by* (2.5).

In the case of symmetric variables X_k the moments $b_{k,3}$ vanish. Therefore:

THEOREM 4. *If* $1 - V_k(x) = V_k(-x)$ *and* (2.10) *holds, then* $\{\phi_n\} \in \mathcal{U}(\mathcal{L})$ *if, and only if, the series* (2.9) *converges (diverges).*

In general we have in the same way the following theorem.

THEOREM 5. *If, for some integer* $\rho > 1$,

$$\lambda_n = O(1/\phi_n^{(\rho+2)/\rho}) \tag{2.12}$$

then $\{\phi_n\} \in \mathcal{U}(\mathcal{L})$ *if, and only if,*

$$\sum_n \frac{\sigma_n^2}{s_n^2} \phi_n \exp\{-(1/2)[\phi_n^2 + q_{n,1}\phi_n^3 + \cdots + q_{n,\rho-1}\phi_n^{\rho+1}]\} \in \mathcal{C}(\mathcal{D}); \tag{2.13}$$

thus the behavior depends only on the first $\rho+1$ *moments of* $\{V_k(x)\}$. *We have in particular*

$$q_{n,1} = \frac{1}{3s_n^3}\sum_{k=1}^n b_{k,3},$$

$$q_{n,2} = \frac{1}{12s_n^4}\sum_{k=1}^n b_{k,4} - \frac{1}{4s_n^4}\sum_{k=1}^n b_{k,2}^2 - \frac{1}{4s_n^6}\left(\sum_{k=1}^n b_{k,3}\right)^2.$$

In §7 it will be proved that the criterion of Theorem 2 can be reformulated

in a way analogous to Kolmogoroff's criterion in the binomial case as follows.

THEOREM 6. *If* $\phi(t)\uparrow$ *and*

$$\lambda_n = O(1/\phi^2(s_n^2)) \tag{2.14}$$

then $\{\phi(s_n^2)\}\in\mathcal{U}(\mathcal{L})$ *if, and only if,*

$$\int^{\infty} \frac{1}{t}\phi(t)e^{-(1/2)\phi^2(t)}dt \in \mathcal{C}(\mathcal{D}). \tag{2.15}$$

The connection between our theorems and the law of the iterated logarithm becomes more apparent in the following theorem.

THEOREM 7. *If either*

$$\lambda_n = O(1/\log_2^{3/2} s_n), \tag{2.16}$$

or $V_k(-x)=1-V_k(x)$ *and*

$$\lambda_n = O(1/\log_2 s_n), \tag{2.17}$$

then

$$\phi_n \equiv \{2\log_2 s_n^2 + 3\log_3 s_n^2 + 2\log_4 s_n^2 + \cdots + 2\log_{p-1} s_n^2 + (2+\delta)\log_p s_n^2\}^{1/2} \in \mathcal{U}(\mathcal{L}) \tag{2.18}$$

if, and only if,

$$\delta > 0 (< 0).$$

This theorem is actually a simple corollary to Theorems 2 and 4. For the proof it suffices to note that, with the definition (2.18), $\phi_n=(2\log_2 s_n)^{1/2}+o(1)$, so that (2.16) and (2.17) are equivalent to (2.8) and (2.10), respectively. Thus the criterion (2.9) is applicable. However,

$$\sum \frac{\sigma_n^2}{s_n^2}\phi_n e^{-(1/2)\phi_n^2} = (2)^{1/2}\sum \frac{\sigma_n^2(1+o(1))}{s_n^2 \log s_n^2 \log_2 s_n^2 \cdots \log_p^{1+\delta/2} s_n^2};$$

now $\sum\sigma_n^2$ diverges, and therefore[14] the above series converges if, and only if, $\delta>0$.

We shall prove the following *sharpening of the notion of lower class*, which holds true in all cases considered in this paper.

---

[14] This follows easily from theorems of Abel-Dini and of Cesàro. See, for example, K. Knopp, *Theory and applications of infinite series*, London-Glasgow, Blackie and Son, 1928, pp. 292 ff. Theorem 7 is, of course, an immediate consequence also of Theorem 6. Specialized to the symmetric binomial case it becomes a consequence of the theorem of Kolmogoroff-Erdös. It should be noted that the terms s_n^2 in (2.18) can be replaced by s_n.

THEOREM 8. *Under the condition* (2.6) *of Theorem* 1, *if* $\{\phi_n\} \in \mathcal{L}$ *then for any couple of constants* $\alpha < \beta$ *and at almost all points, the inequality*

$$s_n(\phi_n + \alpha/\phi_n) < S_n < s_n(\phi_n + \beta/\phi_n) \tag{2.19}$$

will be satisfied for infinitely many n.

Hence: *If* $\{\phi_n\} \in \mathcal{L}$ *also* $\{\phi_n + \alpha/\phi_n\} \in \mathcal{L}$.

Besides the criterion of Theorem 1 we shall prove the following (less handy) criterion:

THEOREM 9. *Under the condition* (2.6) *of Theorem* 1, *the sequence* $\{\phi_n\} \in \mathcal{U}(\mathcal{L})$ *if, and only if, for any sequence of integers* $\{n_k\}$ *for which*

$$s_{n_k}^2(1 + a/\phi_{n_k}^2) < s_{n_{k+1}}^2 < s_{n_k}^2(1 + b/\phi_{n_k}^2), \tag{2.20}$$

where a *and* b *are constants, the series*

$$\sum \Pr\{S_{n_k} > s_{n_k}\phi_{n_k}\} \in \mathcal{C}(\mathcal{D}). \tag{2.21}$$

We mention also the following corollary to Theorem 1, which follows easily by means of (2.4).

THEOREM 10. *Under the condition* (2.6) *of Theorem* 1 *if*

$$\sum_n \frac{\sigma_n^2}{s_n^2}\phi_n e^{-(1/2)\phi_n^2} \in \mathcal{C}, \tag{2.22}$$

the sequence

$$\{\phi_n + \lambda_n\phi_n^2\} \in \mathcal{U}; \tag{2.23}$$

if (2.22) *diverges,*

$$\{\phi_n - \lambda_n\phi_n^2\} \in \mathcal{L}. \tag{2.24}$$

Finally, we give an alternative form for the conditions of the preceding theorems. Its usefulness will be exemplified by the results of the appendix.

THEOREM 11. *The criterion of Theorem* 1 *remains valid if the condition* (2.6) *is replaced by*

$$\lambda_n \leqq 1/200(\log\log s_n)^{1/2}. \tag{2.25}$$

A similar remark applies to Theorems 2–6. *If* (2.25) *is satisfied, the sequence* $\{\phi_n\}$ *belongs to the same class as the sequence* $\{\psi_n\}$ *defined by*

$$\psi_n = \min\{\phi_n, 2(\log\log s_n)^{1/2}\}. \tag{2.26}$$

For a generalization of the preceding results to the case of unbounded variables $\{X_n\}$ the reader is referred to the appendix.

2.3. **Remarks.** The following remarks are either obvious or easily verified. They will not be used in the sequel, and their detailed proof can therefore be left to the reader.

1. Theorem 8 contains a *best result* in two directions. The assertion that $\{\phi_n\}$ and $\{\phi_n+\text{const}/\phi_n\}$ belong to the same class cannot be improved. First, if $\{X_n\}$ is a given sequence and $a_n\to\infty$, it is possible to find a sequence such that $\{\phi_n\}\in\mathcal{L}$ but $\{\phi_n+a_n/\phi_n\}\in\mathcal{U}$ (cf. Erdös [2] in the case of the binomial distribution with equal probabilities). More interesting is that to any given $\{\phi_n\}$ and $a_n\to\infty$ it is possible to find a sequence $\{X_n\}$ such that $\{\phi_n\}\in\mathcal{L}$ but $\{\phi_n+a_n/\phi_n\}\in\mathcal{U}$. The fact that $\{\phi_n\}$ and $\{\phi_n+a/\phi_n\}$ belong to the same class is easily established independently of the previous results. In fact, it is sufficient to notice that $|(\phi_n+a/\phi_n)^2Q_n(\phi_n+a/\phi_n)-\phi_n^2Q_n(\phi_n)|=O(1)$, and this is readily proved using (2.4).

2. If $\lambda_n=O(1/\phi_n^3)$ then $\{\phi_n\}$ is of the same class with respect to the sequence $\{X_n\}$ and $\{-X_n\}$. This is a best result. If $a_n\to\infty$ is an arbitrary sequence it is possible to find a sequence $\{X_n\}$ with $\lambda_n=O(a_n/\phi_n^3)$ such that $\{\phi_n\}\in\mathcal{U}$ for $\{X_n\}$ but $\{\phi_n\}\in\mathcal{L}$ for $\{-X_n\}$.

3. It has been conjectured by P. Lévy [12, p. 266] that Kolmogoroff's criterion (Theorem 6) is applicable for any uniformly bounded sequence $\{X_k\}$. He put the problem of finding the weakest conditions on $\{\lambda_n\}$ under which the theorem would hold. Theorem 6 gives the *best* result. If $\lambda_n\phi_n^3\to\infty$, it is possible to find a sequence $\{X_n\}$ such that the criterion of Theorems 2 and 6 becomes false.

4. It will be noticed that the condition (2.16) is sharper than Kolmogoroff's condition (1.8). However, (2.16) is the *best* condition. If $\lambda_n\log_2^{3/2}s_n\to\infty$ it is possible to construct a sequence $\{X_n\}$ such that (2.18) does not hold.

5. It is interesting to compare Theorem 10 with the classical law of the iterated logarithm, and with our theorems. If $\lambda_n=O(1/\phi_n^3)$ then the sequences $\{\phi_n\pm\lambda_n\phi_n^3\}$ obviously belong to the same class, and Theorem 10 reduces to Theorem 1. However, as the order of magnitude of $\{\lambda_n\}$ increases, Theorem 10 becomes less and less sharp. With the law of the iterated logarithm one knows only that $\lambda_n=o(1/\phi_n)$; hence the assertion relates only to sequences of the form $\{\phi_n\pm\epsilon\phi_n\}$. When $\lambda_n=O(1/\phi_n)$ even this breaks down, which explains the behavior of the counter-example constructed by Marcinkiewicz and Zygmund.

6. It is hardly necessary to point out that the constant 1/200 in Theorem 1 (2.6) is convenient but purely arbitrary. Actually 1/12 would have been sufficient (except for minor complications). There is no best constant, as the radius of convergence of $Q_n(x)$ depends on the sequence $\{X_n\}$.

3. **The sequence $\{n_k\}$.** We proceed first to establish the existence of sequences of the kind occurring in Theorem 9: *If condition* (2.6) *is satisfied then to any constants a, b with*

$$0 < a < (99/100)b, \qquad b > 1/100, \tag{3.1}$$

there exists a sequence $\{n_k\}$ *such that* (2.20) *holds. Obviously* $n_k \to \infty$.

The proof is very simple. Since, by assumption, $s_n \to \infty$ we can choose n_1 arbitrarily and define a sequence $\{n_k\}$ by the recurrence relations

$$s^2_{n_{k+1}} \leqq s^2_{n_k}(1 + b/\phi^2_{n_k}), \tag{3.2}$$

$$s^2_{n_{k+1}+1} > s^2_{n_k}(1 + b/\phi^2_{n_k}). \tag{3.3}$$

Then by (1.5), (2.1), (2.6), and (2.2)

$$s^2_{n_{k+1}+1} = s^2_{n_{k+1}} + \sigma^2_{n_{k+1}+1} \leqq s^2_{n_{k+1}} + 10^{-4}\phi^{-2}_{n_{k+1}+1}s^2_{n_{k+1}+1}$$
$$\leqq s^2_{n_{k+1}} + 10^{-4}\phi^{-2}_{n_k}s^2_{n_{k+1}+1},$$

or, using (3.3) and (3.1),

$$s^2_{n_{k+1}} \geqq s^2_{n_{k+1}+1}(1 - 10^{-4}\phi^{-2}_{n_k}) > s^2_{n_k}(1 + b/\phi^2_{n_k})(1 - 10^{-4}\phi^{-2}_{n_k})$$
$$\geqq s^2_{n_k}(1 + a/\phi^2_{n_k}).$$

In the following $\{n_k\}$ *shall stand for an arbitrary, but fixed, sequence satisfying* (2.20). *We shall write*

$$S_{n_k} = S'_k, \quad s_{n_k} = s'_k, \quad \phi_{n_k} = \phi'_k, \quad Q_{n_k}(x) = Q'_k(x), \quad \lambda_{n_k} = \lambda'_k. \tag{3.4}$$

This will greatly simplify the notation, as we shall require a subsequence of $\{n_k\}$ and, in turn, a subsequence of it.

4. **Lemma** 1. *Suppose that condition* (2.6) *is satisfied and that*

$$\sum \Pr\{S'_k > s'_k\phi'_k\} \in \mathfrak{D}. \tag{4.1}$$

Let $\epsilon > 0$ *be fixed. Then at almost all points the inequality*

$$s_n\phi_n < S_n < s_n(\phi_n + 3\epsilon/\phi_n) \tag{4.2}$$

will be satisfied for infinitely many n.

Before proceeding with the proof let us remark that, putting

$$p_k = (1/\phi'_k)\exp\{-(1/2)\phi'^2_k[1 + Q'_k(\phi'_k)]\}, \tag{4.3}$$

condition (4.1) is equivalent to

$$\sum p_k \in \mathfrak{D}. \tag{4.4}$$

Indeed, by Theorem 1 of L.T.

$$\Pr\{S'_k > s'_k\phi'_k\} < \text{const. } p_k.$$

Moreover, by Theorem 2 of that paper, to any $\epsilon > 0$ there correspond two con-

stants $0<a_1<a_2$, depending only on ϵ, such that

$$a_1 p_k < \Pr\{s_k'\phi_k' < S_k' < s_k'(\phi_k' + \epsilon/\phi_k')\} < a_2 p_k. \tag{4.5}$$

Accordingly, (4.1) is equivalent to the statement that

$$\sum \Pr\{s_k'\phi_k' < S_k' < s_k'(\phi_k' + \epsilon/\phi_k')\} \in \mathfrak{D}. \tag{4.6}$$

4.1. For the proof of Lemma 1 we require a sequence $\{k_\nu\}$ defined by the recurrence relations

$$s'_{k_{\nu+1}} \leqq M s'_{k_\nu}\phi'_{k_\nu}, \qquad s'_{k_{\nu+1}+1} > M s'_{k_\nu}\phi'_{k_\nu}, \tag{4.7}$$

where $M>1$ is a constant to be determined later; k_1 can be chosen arbitrarily. It is readily seen that

$$(2/(2+b))Ms'_{k_\nu}\phi'_{k_\nu} < s'_{k_{\nu+1}} \leqq Ms'_{k_\nu}\phi'_{k_\nu} \tag{4.8}$$

and

$$0 < k_{\nu+1} - k_\nu \leqq (M^2/a)\phi'^4_{k_{\nu+1}}. \tag{4.9}$$

In fact, using (4.7), (2.20), (3.4), and (2.2) it follows that

$$M^2 s'^2_{k_\nu}\phi'^2_{k_\nu} < s'^2_{k_\nu+1} \leqq s'^2_{k_\nu+1}(1 + b/\phi'^2_{k_\nu+1}) \leqq (1 + b/4)s'^2_{k_\nu+1},$$

which obviously implies (4.8). In the same way

$$\phi'^2_{k_\nu} \geqq \frac{1}{M^2}\frac{s'^2_{k_\nu+1}}{s'^2_{k_\nu}} \geqq \frac{1}{M^2}\cdot\prod_{n=k_\nu}^{k_{\nu+1}-1}\left(1 + \frac{a}{\phi'^2_n}\right) \geqq \frac{1}{M^2}\left(1 + \frac{a}{\phi'^2_{k_\nu+1}}\right)^{(k_{\nu+1}-k_\nu)}$$

$$\geqq \frac{1}{M^2}\frac{a}{\phi'^2_{k_\nu+1}}(k_{\nu+1} - k_\nu),$$

which proves (4.9).

4.2. For any $i<j$ let us define

$$S'_{i,j} = S'_j - S'_i = \sum_{n=n_i+1}^{n_j} X_n. \tag{4.10}$$

Then

$$F'_{i,j}(x) = \Pr\{S'_{i,j} \leqq x\} \tag{4.11}$$

is the distribution function of $S'_{i,j}$, and obviously

$$\int_{-\infty}^{+\infty} x dF'_{i,j}(x) = 0 \tag{4.12}$$

and

$$s_{i,j}'^2 \equiv \int_{-\infty}^{+\infty} x^2 dF_{i,j}'(x) = s_j'^2 - s_i'^2. \tag{4.13}$$

Now denote by $A_{\nu,m}$ $(m \geqq k_\nu)$ the set

$$A_{\nu,m} = \underset{p}{E}\{s_m'(\phi_m' + \epsilon/\phi_m') < S_{k_{\nu-1},m}' < s_m'(\phi_m' + 2\epsilon/\phi_m')\}, \tag{4.14}$$

where ϵ is an arbitrary, but fixed, number with

$$0 < \epsilon < 1. \tag{4.15}$$

We propose now to show: *if*

$$k_\nu \leqq m < k_{\nu+1}, \tag{4.16}$$

$$\phi_{k_{\nu+1}}' < 2\phi_{k_{\nu-1}}', \tag{4.17}$$

and if

$$(2 + b)^2/M^2 < \epsilon/4, \tag{4.18}$$

then there exists a constant $\eta > 0$ (depending only on ϵ) such that

$$\Pr(A_{\nu,m}) > \eta p_m \tag{4.19}$$

for all ν sufficiently large.

Proof. Since the sequence s_k' is not decreasing and $m > k_\nu$ we obtain, using (4.16), (4.8), (4.17), and (4.18),

$$s_{k_{\nu-1},m}'^2 = s_m'^2 - s_{k_{\nu-1}}'^2 \geqq s_m'^2\{1 - s_{k_{\nu-1}}'^2/s_{k_\nu}'^2\} \geqq s_m'^2\{1 - (2 + b)^2/(4M^2\phi_{k_{\nu-1}}'^2)\}$$
$$\geqq s_m'^2\{1 - (2 + b)^2/(M^2\phi_m'^2)\} \geqq s_m'^2\{1 - \epsilon/(4\phi_m'^2)\},$$

whence

$$s_m'(1 - \epsilon/(4\phi_m'^2)) \leqq s_{k_{\nu-1},m}' \leqq s_m'. \tag{4.20}$$

Accordingly, putting

$$A_{\nu,m}^* = \underset{p}{E}\{s_{k_{\nu-1},m}'(\phi_m' + 3\epsilon/(2\phi_m')) \leqq S_{k_{\nu-1},m}' \leqq s_{k_{\nu-1},m}'(\phi_m' + 2\epsilon/\phi_m')\} \tag{4.21}$$

it is readily seen that

$$A_{\nu,m}^* \subset A_{\nu,m}. \tag{4.22}$$

To evaluate $\Pr(A_{\nu,m}^*)$ we can apply Theorems 1 and 2 of L.T. Indeed, $S_{k_{\nu-1},m}'$ is a sum of independent random variables X_k; moreover, it is easily seen (using (4.20), (4.15), (2.2) and (2.6)) that

$$\text{l.u.b.} \mid X_k \mid < \lambda_{n_m} s_m'(17/16)\lambda_{n_m} s_{k_{\nu-1},m}' < (1/100)s_{k_{\nu-1},m}',$$

so that Theorem 2 is applicable. It follows immediately that

$$\Pr(A_{\nu,m}) \geqq \Pr(A^*_{\nu,m}) \geqq \eta p_m.$$

4.3. Now put

$$A_\nu = \sum_{m=k_\nu}^{k_{\nu+1}-1} A_{\nu,m} \tag{4.23}$$

and

$$P_\nu = \sum_{m=k_\nu}^{k_{\nu+1}-1} p_m. \tag{4.24}$$

We shall show: *If* (4.17) *and* (4.19) *hold, then there exists a constant* $\eta_1>0$ *such that*

$$\Pr(A_\nu) > \eta_1 P_\nu, \tag{4.25}$$

for all ν *sufficiently large.*

Proof. Let N be a fixed integer and consider the sets

$$A^{(N)}_{\nu,m} = A_{\nu,m} - A_{\nu,m} \sum_{\rho=m+N}^{k_{\nu+1}-1} A_{\nu,\rho}. \tag{4.26}$$

Obviously

$$A_\nu = \sum_{m=k_\nu}^{k_{\nu+1}-1} A^{(N)}_{\nu,m}. \tag{4.27}$$

On the other hand, it follows from (4.26) that the intersection of two sets $A^{(N)}_{\nu,m_1}$ and $A^{(N)}_{\nu,m_2}$ will be empty unless $|m_1-m_2|<N$. Any point of A_ν is therefore contained in at most N distinct sets $A^{(N)}_{\nu,m}$ figuring in (4.27) so that

$$\Pr(A_\nu) \geqq \frac{1}{N} \sum_{m=k_\nu}^{k_{\nu+1}-1} \Pr(A^{(N)}_{\nu,m}). \tag{4.28}$$

Again it follows from the definition (4.14) of $A_{\nu,m}$ that for any point of $A_{\nu,m}A_{\nu,\rho}$ $(\rho>m)$

$$S'_{m,\rho} > s'_\rho(\phi'_\rho + \epsilon/\phi'_\rho) - s'_m(\phi'_m + 2\epsilon/\phi'_m) > (s'_\rho - s'_m)\phi'_\rho - s'_m(2\epsilon/\phi'_m). \tag{4.29}$$

In other words

$$A_{\nu,m}A_{\nu,\rho} \subset B_{m,\rho}, \tag{4.30}$$

where

$$B_{m,\rho} = E_\rho\{S'_{m,\rho} > (s'_\rho - s'_m)\phi'_\rho - s'_m(2\epsilon/\phi'_m)\}. \tag{4.31}$$

However, the sets $A_{\nu,m}$ and $B_{m,\rho}$ are independent (see the footnote 5 on p. 375), so that

$$P(A_{\nu,m}B_{m,\rho}) = P(A_{\nu,m})P(B_{m,\rho}). \tag{4.32}$$

Comparing (4.26), (4.28), (4.30), (4.32), and (4.19) we see that

$$\begin{aligned}\Pr(A_\nu) &\geqq \frac{1}{N}\sum_{m=k_\nu}^{k_{\nu+1}-1}\Pr\Big\{A_{\nu,m} - \sum_{\rho=m+N}^{k_{\nu+1}-1} A_{\nu,m}B_{m,\rho}\Big\} \\ &\geqq \frac{1}{N}\sum_{m=k_\nu}^{k_{\nu+1}-1}\Pr(A_{\nu,m})\Big\{1 - \sum_{\rho=m+N}^{k_{\nu+1}-1}\Pr(B_{m,\rho})\Big\} \\ &\geqq \frac{\eta}{N}\sum_{m=k_\nu}^{k_{\nu+1}-1} p_m\Big\{1 - \sum_{\rho=m+N}^{k_{\nu+1}-1}\Pr(B_{m,\rho})\Big\}.\end{aligned} \tag{4.33}$$

We turn to the evaluation of $\Pr(B_{m,\rho})$. For that purpose we write the inequality in (4.31) in the form

$$S'_{m,\rho} > s'_{m,\rho}\phi^*_\rho \tag{4.34}$$

where

$$\phi^*_\rho = \frac{s'_\rho - s'_m}{s'_{m,\rho}}\,\phi'_\rho - \frac{s'_m}{s'_{m,\rho}}\cdot\frac{2\epsilon}{\phi'_m}. \tag{4.35}$$

To evaluate the probability of (4.34) we may again apply Theorem 1 of L.T., provided that all the X_n with $n<n_\rho$ have an upper bound not exceeding, say, $(1/200\phi^*_\rho)s'_{m,\rho}$. Now for $n<n_\rho$

$$|X_n| < \lambda'_\rho s'_\rho \leqq \frac{1}{200\phi'_\rho}\,\frac{s'_\rho}{s'_{m,\rho}}\,s'_{m,\rho}$$

and by (4.35)

$$\frac{\phi^*_\rho}{\phi'_\rho}\,\frac{s'_\rho}{s'_{m,\rho}} < \frac{s'_\rho(s'_\rho - s'_m)}{s'^2_{m,\rho}} = \frac{s'_\rho}{s'_m + s'_\rho} < 1.$$

Theorem 1 of L.T. is therefore applicable and

$$\begin{aligned}\Pr(B_{m,\rho}) &= \Pr\{S'_{m,\rho} > s'_{m,\rho}\phi^*_\rho\} \\ &< \eta_2\frac{1}{\phi^*_\rho}\exp\{-(1/2)\phi^{*2}_\rho[1 + Q^*_{n_\rho - n_m}(\phi^*_\rho)]\}\end{aligned} \tag{4.36}$$

provided, say, $\phi^*_\rho > 2$. Using the estimates (2.4) for the coefficients of Q^* it is readily seen that

$$\Pr(B_{m,\rho}) \leqq \eta_2 e^{(-2/5)\phi^{*2}_\rho}. \tag{4.37}$$

Now by (4.35)

$$\phi_\rho^* = \frac{s_\rho' - s_m'}{(s_\rho'^2 - s_m'^2)^{1/2}}\,\phi_\rho' - \frac{s_m'}{(s_\rho'^2 - s_m'^2)^{1/2}}\,\frac{2\epsilon}{\phi_m'}$$

$$= \frac{(s_\rho'^2 - s_m'^2)^{1/2}}{s_\rho' + s_m'}\,\phi_\rho' - \frac{s_m'}{(s_\rho'^2 - s_m'^2)^{1/2}}\,\frac{2\epsilon}{\phi_m'} \tag{4.38}$$

$$\geqq \frac{1}{2}\left(1 - \frac{s_m'^2}{s_\rho'^2}\right)^{1/2}\phi_\rho' - \frac{1}{(s_\rho'^2/s_m'^2 - 1)^{1/2}}\,\frac{2\epsilon}{\phi_m'}.$$

However, by means of (2.20), (3.4), (4.16) and (4.17) we deduce that for $\rho < k_{\nu+1}$

$$\left(\frac{s_\rho'}{s_m'}\right)^2 \geqq \prod_{n=m}^{\rho-1}\left(1 + \frac{a}{\phi_n'^2}\right) \geqq \left(1 + \frac{a}{4\phi_m'^2}\right)^{\rho-m} \geqq 1 + \frac{a(\rho-m)}{4\phi_m'^2}. \tag{4.39}$$

This inequality will be applied in three ways.

(i) Since $\rho \geqq m+N$ it follows that

$$s_\rho'^2/s_m'^2 \geqq 1 + aN/(4\phi_\rho'^2);$$

using (4.17) it is therefore seen that the last term of the right-hand member in (4.38) is small if N is large. Hence (4.38) implies, for N sufficiently large,

$$\phi_\rho^* \geqq (1/2)(1 - s_m'^2/s_\rho'^2)^{1/2}\phi_\rho' - a^{1/2}/10. \tag{4.40}$$

(ii) Consider now ρ with

$$a(\rho - m) \leqq \phi_m'^2. \tag{4.41}$$

Combining (4.39)–(4.41) we obtain

$$\phi_\rho^* \geqq \frac{1}{2}\left\{1 - \frac{1}{1 + a(\rho-m)/(4\phi_m'^2)}\right\}^{1/2}\phi_\rho' - \frac{a^{1/2}}{10} \geqq \frac{a^{1/2}}{10}(\rho - m)^{1/2}. \tag{4.42}$$

Comparing (4.37) with (4.42) it is readily seen that, for N sufficiently large,

$$\sum_{\rho=m+N}^{\min\{k_{\nu+1}-1,\, m+[a^{-1}\phi_m'^2]\}} \Pr(B_{m,\rho}) \leqq \eta_2 \sum_{\rho=m+N}^{\infty} e^{-(a/250)(\rho-m)} < \frac{1}{3}. \tag{4.43}$$

(iii) Finally, consider ρ with

$$a(\rho - m) \geqq \phi_m'^2. \tag{4.44}$$

For these we obtain from (4.39) $s_\rho'^2/s_m'^2 \geqq 5/4$ and from (4.40)

$$\phi_\rho^* \geqq (1/10)(\phi_\rho' - a^{1/2}); \tag{4.45}$$

since $k_\nu < \rho < k_{\nu+1}$ and since $\phi_n' \to \infty$, it follows from (4.45) and (4.17) that for ν sufficiently large

$$\phi_\rho^* \geqq (1/30)\phi'_{k_{\nu+1}}. \tag{4.46}$$

Combining (4.37) and (4.46) shows that

$$\begin{aligned}\sum_{\rho=m+[\phi_m'^2/\alpha]}^{k_{\nu+1}-1} \Pr(B_{m,\rho}) &< \eta_2 \sum_{\rho=m+[\phi_m'^2/\alpha]}^{k_{\nu+1}-1} e^{-(1/2500)\phi'_{k_{\nu+1}}} \\ &\leqq \eta_2(k_{\nu+1} - k_\nu)e^{-(1/2500)\phi'_{k_{\nu+1}}}.\end{aligned} \tag{4.47}$$

However, (4.9) implies that the last expression tends to zero as $\nu \to \infty$. Accordingly, for ν sufficiently large,

$$\sum_{\rho=m+[\phi_m'^2/\alpha]}^{k_{\nu+1}-1} \Pr(B_{m,\rho}) < \frac{1}{3}. \tag{4.48}$$

Substituting for $\Pr(B_{m,\rho})$ from (4.43) and (4.48) in (4.33) we obtain

$$\Pr(A_\nu) \geqq \frac{\eta}{3N} \sum_{m=k_\nu}^{k_{\nu+1}-1} p_m = \frac{\eta}{3N} P_\nu,$$

which proves the assertion (4.25).

4.4. We propose now to show that almost all points belong to infinitely many A_ν, or, in other words, that

$$\Pr\left(\sum_{\nu=\bar{\nu}}^{\infty} A_\nu\right) = 1, \tag{4.49}$$

for any $\bar{\nu}$. For that purpose we note that the set A_ν depends only on the variables X_n with $n_{k_{\nu-1}} \leqq n < n_{k_{\nu+1}}$ (cf. (4.14) and (4.23)). Accordingly, the sets $A_{2\nu}$ are mutually independent, and

$$\Pr\left(\sum_{\nu=\bar{\nu}}^{\infty} A_\nu\right) \geqq \Pr\left(\sum_{\nu=\bar{\nu}}^{\infty} A_{2\nu}\right) = 1 - \prod_{\nu=\bar{\nu}}^{\infty}(1 - \Pr(A_{2\nu})), \tag{4.50}$$

and a similar relationship holds for odd subscripts. Now it would be trivial to deduce (4.49) from (4.50), if we could use (4.25). However, the last inequality has been proved only under the assumption that (4.17) holds. We shall show that the ν's for which (4.17) does not hold are of no influence.

Let[18] $\{\nu(r)\}$, $r = 1, \cdots$, denote those integers for which

$$\phi'_{k_{\nu,r}+1} \geqq 2\phi'_{k_{\nu(r)-1}}. \tag{4.51}$$

We shall prove that

$$\sum_r P_{\nu(r)} \in \mathcal{C}. \tag{4.52}$$

[18] For simplicity in print the notation $\nu(r)$ has been introduced instead of the usual ν_r.

In fact, it follows readily from (2.6), (2.4) and (4.3) that $p_k \leqq (1/\phi_k')e^{(-2/\delta)\phi_k'^2}$, and therefore

$$p_k \leqq (1/\phi_k'^4)e^{-\phi_k'} \tag{4.53}$$

for k sufficiently large. We find by means of (4.8) and (2.20)

$$\phi_{k_\nu}'^2 \geqq \frac{1}{M^2}\frac{s_{k_{\nu+1}}'^2}{s_{k_\nu}'^2} \geqq \frac{1}{M^2}\prod_{m=k_\nu}^{k_{\nu+1}-1}\left(1+\frac{a}{\phi_m'^2}\right) \geqq \frac{a}{M^2}\sum_{m=k_\nu}^{k_{\nu+1}-1}\frac{1}{\phi_m'^2}.$$

Thus by (4.53) and (2.2)

$$P_\nu = \sum_{m=k_\nu}^{k_{\nu+1}-1} p_m \leqq \frac{e^{-\phi_{k_\nu}'}}{\phi_{k_\nu}'^2}\sum_{m=k_\nu}^{k_{\nu+1}-1}\frac{1}{\phi_m'^2} \leqq \frac{M^2}{a}e^{-\phi_{k_\nu}'}$$

or

$$\sum P_{\nu(r)} \leqq \frac{M^2}{a}\sum_r e^{-\phi_{k_{\nu(r)}}'}; \tag{4.54}$$

but (4.51) implies that the sequence $\{\phi_{k_{\nu(r)}}'\}$ increases so rapidly that the series (4.54) converges. Accordingly the series $\sum_{\nu\neq\nu(r)}P_\nu$ diverges, and without loss of generality we may suppose that $\sum_{2\nu\neq\nu(r)}P_{2\nu}$ diverges. It follows from the definition (4.51) that for $\nu\neq\nu(r)$, (4.17) holds. Accordingly, for these ν the inequality (4.25) holds, and therefore

$$\begin{aligned}\Pr\left(\sum_{\nu\geqq\bar{\nu},\,2\nu\neq\nu(r)} A_{2\nu}\right) &= 1 - \prod_{\nu\geqq\bar{\nu},\,2\nu\neq\nu(r)}{}^{*}\,(1-\Pr(A_{2\nu})) \\ &\geqq 1 - \prod_{\nu\geqq\bar{\nu},\,2\nu\neq\nu(r)}(1-\eta_1 P_\nu) = 1.\end{aligned} \tag{4.55}$$

4.5. The proof of Lemma 1 is now easily accomplished. We know that almost all points belong to infinitely many sets $A_{2\nu}$ with $2\nu\neq\nu(r)$, that is to say, such that

$$\phi_{k_{2\nu+1}}' < 2\phi_{k_{2\nu-1}}'. \tag{4.56}$$

Denote then by $C_{2\nu}$ the set

$$C_{2\nu} = \underset{p}{E}\{-(1/\delta)s_{k_{2\nu-1}}' < S_{k_{2\nu-1}}' < (1/\delta)s_{k_{2\nu+1}}'\}, \tag{4.57}$$

where $\delta>0$ is arbitrary. At any point of the intersection $A_{2\nu,m}C_{2\nu}$ we have simultaneously the inequality figuring in (4.57) and

$$s_m'(\phi_m' + \epsilon/\phi_m') < S_{k_{2\nu-1},m}' < s_m'(\phi_m' + 2\epsilon/\phi_m'). \tag{4.58}$$

Adding these inequalities we find that in $A_{2\nu,m}C_{2\nu}$

$$s_m'\phi_m' + s_m'\epsilon/\phi_m' - (1/\delta)s_{k_{2\nu-1}}' < S_m' < s_m'\phi_m' + 2\epsilon\, s_m'/\phi_m' + (1/\delta)s_{k_{2\nu+1}}'. \tag{4.59}$$

However, for $k_{2\nu} \leqq m < k_{2\nu+1}$ we have, using (4.8) and (4.56),

$$\frac{s'_{k_{2\nu}-1}}{s'_m} \leqq \frac{s'_{k_{2\nu}-1}}{s'_{k_{2\nu}}} \leqq \frac{2+b}{2M}\,\frac{1}{\phi'_{k_{2\nu}-1}} \leqq \frac{2+b}{M}\,\frac{1}{\phi'_m};$$

thus, if the constant M is chosen so that

$$(2+b)/M < \epsilon\delta \tag{4.60}$$

we obtain from (4,59) that at any point of $A_{2\nu,m}C_{2\nu}$

$$s'_m\phi'_m < S'_m < s'_m(\phi'_m + 3\epsilon/\phi'_m). \tag{4.61}$$

Denote the set of points for which (4.61) holds by Λ_m. Then $\Lambda_m \supset A_{2\nu,m}C_{2\nu}$, and therefore (always for $2\nu \neq \nu(r)$)

$$\Pr\Big(\sum_{m=k_{2\nu}}^{k_{2\nu+1}-1} \Lambda_m\Big) \geqq \Pr\,(A_{2\nu}C_{2\nu})$$

or

$$\Pr\Big(\sum_{m \geqq k_\nu'} \Lambda_m\Big) \geqq \Pr\Big(\sum_{\nu \geqq \nu,\, 2\nu \neq \nu(r)} A_{2\nu}C_{2\nu}\Big). \tag{4.62}$$

Again, it follows from the definition (4.57) that (Tschebycheff's inequality)

$$\begin{aligned}\Pr\,(C_{2\nu}) &= \int_{|x|<(1/\delta)s'_{k_{2\nu-1}}} dF'_{k_{2\nu-1}} = 1 - \int_{|x|\geqq(1/\delta)s'_{k_{2\nu-1}}} dF'_{k_{2\nu-1}} \\ &\geqq 1 - (\delta^2/s'^2_{k_{2\nu-1}}) \int_{|x|\geqq(1/\delta)s'_{k_{2\nu-1}}} x^2\,dF'_{k_{2\nu-1}} \geqq 1 - \delta^2.\end{aligned} \tag{4.63}$$

On the other hand, the sets $A_{2\nu}$ and $C_{2\rho}$ are independent if $\rho \geqq \nu$ (cf. footnote 5 on p. 375). We have therefore

$$\begin{aligned}\Pr\Big(\sum_{\nu \geqq \nu,\, 2\nu \neq \nu(r)} A_{2\nu}C_{2\nu}\Big) &= \sum_{\nu \geqq \nu,\, 2\nu \neq \nu(r)} \Pr\Big\{A_{2\nu}C_{2\nu} - A_{2\nu}C_{2\nu}\sum_{\rho>\nu,\, 2\rho \neq \nu(r)} A_{2\rho}C_{2\rho}\Big\} \\ &\geqq \sum_{\nu \geqq \nu,\, 2\nu \neq \nu(r)} \Pr\Big\{C_{2\nu}\Big(A_{2\nu} - A_{2\nu}\sum_{\rho>\nu,\, 2\rho \neq \nu(r)} A_{2\rho}\Big)\Big\} \\ &= \sum_{\nu \geqq \nu,\, 2\nu \neq \nu(r)} \Pr\,(C_{2\nu})\,\Pr\Big(A_{2\nu} - A_{2\nu}\sum_{\rho>\nu,\, 2\rho \neq \nu(r)} A_{2\rho}\Big) \\ &\geqq (1-\delta^2)\,\Pr\Big(\sum_{\nu \geqq \nu,\, 2\nu \neq \nu(r)} A_{2\nu}\Big) = (1-\delta^2)\end{aligned}$$

(by (4.55)). This proves that all points, with the possible exception of a set of measure δ^2, belong to infinitely many sets Λ_m; at such points the inequality

(4.61) is satisfied for infinitely many values m. Since δ is arbitrary, and the definition of Λ_m is independent of δ, it follows that at almost all points the inequality (4.61) will be satisfied for infinitely many m. This proves Lemma 1.

5. **Lemma 2.** *Suppose that condition* (2.6) *is satisfied and that*

$$\sum \Pr\{S_k' > s_k'\phi_k'\} \in \mathcal{C}. \tag{5.1}$$

Then

$$\{\phi_n\} \in \mathcal{L}. \tag{5.2}$$

As before (cf. §4) we see that condition (5.1) is equivalent to

$$\sum p_k \in \mathcal{C} \tag{5.3}$$

where p_k is defined by (4.3). We shall require three sets:

$$A_n = \underset{p}{E}\{S_n > s_n\phi_n\}, \tag{5.4}$$

$$B_{n,k} = \underset{p}{E}\{S_{n,n_k} > -2s_{n,n_k}\}, \qquad n_{k-1} < n \leqq n_k, \tag{5.5}$$

$$C_k = \underset{p}{E}\{S_{n_k} > s_{n_k}(\phi_{k-1}' - (2+2b)/\phi_{k-1}')\} \tag{5.6}$$

(for the definition of $S_{n,m}$ and $s_{n,m}$ cf. (4.10) and (4.13)). It is sufficient to prove that

$$P\left(\sum_{n=N}^{\infty} A_n\right) \leqq \epsilon \tag{5.7}$$

for N sufficiently large.

Again we deduce from (2.20) for $n_{k-1} < n \leqq n_k$

$$s_n^2 \geqq s_{n_{k-1}}^2 \geqq s_{n_k}^2/(1 + b/\phi_{n_{k-1}}^2) \geqq s_{n_k}^2(1 - b/\phi_{n_{k-1}}^2); \tag{5.8}$$

on the other hand by definition and by (5.8)

$$s_{n,n_k}^2 = s_{n_k}^2 - s_n^2 \leqq (b/\phi_{n_{k-1}}^2)s_{n_k}^2.$$

Adding the inequalities in (5.4) and (5.5) and using again (5.8) we see therefore that at all points of the intersection $A_nB_{n,k}$

$$\begin{aligned} S_{n_k} \equiv S_n + S_{n,n_k} &> s_{n_k}(\phi_n - b\phi_n/\phi_{n_{k-1}}^2 - 2b^{1/2}/\phi_{n_{k-1}}) \\ &= s_{n_k}(\phi_{n_{k-1}} + [\phi_n - \phi_{n_{k-1}}][1 - b/\phi_{n_{k-1}}] - [b + 2b^{1/2}]/\phi_{n_{k-1}}) \\ &> s_{n_k}(\phi_{n_{k-1}} - (2+2b)/\phi_{n_{k-1}}), \end{aligned} \tag{5.9}$$

provided $\phi_{n_{k-1}}^2 > b$. Comparing (5.6) with (5.9) it is seen that

$$C_k \supset A_nB_{n,k}, \qquad n_{k-1} < n \leqq n_k. \tag{5.10}$$

Furthermore, using Tschebycheff's inequality (cf. (4.63)), we obtain

$$\Pr(B_{n,k}) \geqq 3/4. \tag{5.11}$$

Now it should be noticed that the sets $B_{n,k}$ and A_m are independent if $m \leqq n$. Therefore

$$\begin{aligned}
\Pr(C_k) &\geqq \Pr\Big(\sum_{n=n_{k-1}+1}^{n_k} A_n B_{n,k}\Big) \\
&= \sum_{n=n_{k-1}+1}^{n_k} \Pr\Big\{A_n B_{n,k} - A_n B_{n,k} \sum_{m=n_{k-1}+1}^{n} A_m B_{m,k}\Big\} \\
&\geqq \sum_{n=n_{k-1}+1}^{n_k} \Pr\Big\{A_n B_{n,k} - A_n B_{n,k} \sum_{m=n_{k-1}+1}^{n} A_m\Big\} \\
&= \sum_{n=n_{k-1}+1}^{n_k} \Pr(B_{n,k}) \Pr\Big(A_n - A_n \sum_{m=n_{k-1}+1}^{n} A_m\Big) \\
&\geqq \frac{3}{4} \Pr\Big(\sum_{n=n_{k-1}+1}^{n_k} A_n\Big),
\end{aligned}$$

or

$$\Pr\Big(\sum_{n=n_{k-1}+1}^{n_k} A_n\Big) \leqq \frac{4}{3} \Pr(C_k). \tag{5.12}$$

However, Theorem 2 of L.T. asserts that $\Pr(C_k) < \eta_2 p_{k-1}$, where η_2 is a constant independent of k. Accordingly

$$\Pr\Big(\sum_{n=n_N}^{\infty} A_n\Big) \leqq \sum_{k=N-1}^{\infty} \Pr\Big(\sum_{n=n_{k-1}+1}^{n_k} A_n\Big) \leqq \frac{4}{3}\eta_2 \sum_{k=N-1}^{\infty} p_{k-1},$$

which proves (5.7).

6. **Lemma 3.** *Suppose that condition* (2.6) *is satisfied, and that the series* (2.7) *diverges (converges). Then*

$$\sum_k \frac{1}{\phi_{n_k}} e^{-(1/2)\phi_{n_k}^2(1+Q_{n_k}(\phi_{n_k}))} \tag{6.1}$$

diverges (converges) for any sequence $\{n_k\}$ *satisfying* (2.20).

REMARK. It was pointed out earlier (cf. the introductory remarks to §§4 and 5) that (6.1) converges or diverges simultaneously with

$$\sum_k \Pr\{S_{n_k} > s_{n_k}\phi_{n_k}\}. \tag{6.2}$$

Thus Lemmas 1 and 2 together are equivalent to Theorem 9 with the sharper form for upper functions provided by Theorem 8. The three lemmas together prove Theorem 1 and Theorem 8. From the previous lemmas it follows that the behavior of (6.1) is necessarily the same for *all* sequences $\{n_k\}$ satisfying (2.20). However, it should be noted that the following proof is absolutely independent of the previously established results.

Proof. As before, let $\{n_k\}$ stand for an arbitrary but fixed sequence satisfying (2.20). Since $\sigma_n^2 \leqq \lambda_n^2 s_n^2 = o(s_n^2)$, we shall have, for sufficiently large[16] n

$$2\sigma_n^2/s_n^2 \geqq \log s_n^2/s_{n-1}^2 = -\log(1-\sigma_n^2/s_n^2) \geqq \sigma_n^2/s_n^2, \tag{6.3}$$

and hence

$$\sum_{n=n_k+1}^{n_{k+1}} \sigma_n^2/s_n^2 \leqq \log s_{n_{k+1}}^2/s_{n_k}^2 \leqq \log(1+b/\phi_{n_k}^2) \leqq b/\phi_{n_k}^2, \tag{6.4}$$

and similarly

$$\sum_{n=n_k+1}^{n_{k+1}} \sigma_n^2/s_n^2 \geqq (1/2)\log(1+a/\phi_{n_k}^2) \geqq a/(4\phi_{n_k}^2). \tag{6.5}$$

Suppose that the series (2.7) diverges, and define the sequence $\{m_k\}$ by

$$\phi_{m_k}\exp\{-(1/2)\phi_{m_k}[1+Q_{m_k}(\phi_{m_k})]\} = \max_{n_k<n\leqq n_{k+1}} \phi_n \exp\{-(1/2)\phi_n^2[1+Q_n(\phi_n)]\}. \tag{6.6}$$

Then

$$n_k < m_k \leqq n_{k+1} \tag{6.7}$$

and by (6.4)

$$\sum_n \frac{\sigma_n^2}{s_n^2}\phi_n e^{-(1/2)\phi_n^2(1+Q_n(\phi_n))} \leqq \sum_k \phi_{m_k} e^{-(1/2)\phi_{m_k}^2(1+Q_{m_k}(\phi_{m_k}))} \sum_{n=n_k+1}^{n_{k+1}} \frac{\sigma_n^2}{s_n^2}$$
$$\leqq b\sum_k \frac{\phi_{m_k}}{\phi_{n_k}^2} e^{-(1/2)\phi_{m_k}^2(1+Q_{m_k}(\phi_{m_k}))}. \tag{6.8}$$

Unfortunately, the ratio ϕ_{m_k}/ϕ_{n_k} is not necessarily bounded. Still, we shall show that those k, for which ϕ_{m_k}/ϕ_{n_k} is not near unity, can be discarded. For that purpose, we make the following simple remark, which will be useful in the sequel: *If for a sequence*[15] $k(1)<k(2)<k(3)<\cdots$

$$\phi_{n_{k(\nu)+1}} > \phi_{n_{k(\nu)}} + 2/\phi_{n_{k(\nu)}} \tag{6.9}$$

then

[16] The argument leading from (6.3) to (6.4) and (6.5) is due to Cesàro; cf. Knopp, loc. cit. p. 292.

$$\sum_r e^{-(1/4)\phi^2_{n_{k(r)}}} \tag{6.10}$$

converges. In fact, the terms of (6.10) decrease geometrically since $(1/4)\phi^2_{n_{k(r+1)}} \geqq (1/4)\phi^2_{n_{k(r)}} + 1$. Now it follows from (2.6) and (6.7) that, at least for k sufficiently large,

$$\phi_{m_k} e^{-(1/2)\phi^2_{m_k}(1+Q_{m_k}(\phi_{m_k}))} > e^{-(1/4)\phi^2_{m_k}} \geqq e^{-(1/4)\phi^2_{n_k}}, \tag{6.11}$$

so that by (6.10)

$$\sum_r \phi_{m_{k(r)}} e^{-(1/2)\phi^2_{m_{k(r)}}(1+Q_{m_{k(r)}}(\phi_{m_{k(r)}}))} < C \neq \infty. \tag{6.12}$$

Again, if $k \neq k(1), k(2), \cdots$ we have

$$\phi_{m_k} \leqq \phi_{n_{k+1}} < 2\phi_{n_k}. \tag{6.13}$$

Using (6.12) and (6.13) we have

$$\sum_k \frac{\phi_{m_k}}{\phi^2_{n_k}} e^{-(1/2)\phi^2_{m_k}(1+Q_{m_k}(\phi_{m_k}))} = \sum_{k \neq k(r)} + \sum_{k(r)} \leqq 4 \sum_{k \neq k(r)} \frac{1}{\phi_{m_k}} e^{-(1/2)\phi^2_{m_k}(1+Q_{m_k}(\phi_{m_k}))} + C;$$

therefore it follows from (6.8) that

$$\sum_{k \neq k(r)} \frac{1}{\phi_{m_k}} e^{-(1/2)\phi^2_{m_k}(1+Q_{m_k}(\phi_{m_k}))} \tag{6.14}$$

diverges.

It will be noticed that the sequence $\{m_k\}$ does not necessarily satisfy the condition (2.20). We shall show that the divergence of (6.14) implies that of (6.1). It suffices obviously to show that if $k \neq k(1), k(2), \cdots$

$$\frac{1}{\phi_{m_k}} e^{-(1/2)\phi^2_{m_k}(1+Q_{m_k}(\phi_{m_k}))} \leqq C \frac{1}{\phi_{n_k}} e^{-(1/2)\phi^2_{n_k}(1+Q_{n_k}(\phi_{n_k}))}, \tag{6.15}$$

where C is a constant. Now $\phi_{m_k} \geqq \phi_{n_k}$. As for the exponential terms we have

$$\begin{aligned} |\phi^2_{m_k}(1+Q_{m_k}(\phi_{m_k})) - \phi^2_{n_k}(1+Q_{n_k}(\phi_{n_k}))| &\leqq (\phi^2_{m_k} - \phi^2_{n_k})(1+|Q_{n_k}(\phi_{n_k})|) \\ &+ \phi^2_{m_k}|Q_{m_k}(\phi_{n_k}) - Q_{n_k}(\phi_{n_k})| + \phi^2_{m_k}|Q_{m_k}(\phi_{m_k}) - Q_{m_k}(\phi_{n_k})|. \end{aligned} \tag{6.16}$$

Now $|Q_n(\phi_n)| < 1/10$ by (2.6) and (2.4). Moreover, since $k \neq k(1), k(2), \cdots$, we have (cf. (6.9))

$$\phi^2_{m_k} - \phi^2_{n_k} \leqq \phi^2_{n_{k+1}} - \phi^2_{n_k} \leqq 4 + 4/\phi^2_{n_k} < 5. \tag{6.17}$$

Thus the first term in (6.16) is less than 6.

Next, using Theorem 1, formula (2.8), of L.T. and (2.20) we find

$$\phi^2_{m_k}|Q_{m_k}(\phi_{n_k} - Q_{n_k}(\phi_{n_k})| \leqq 2\phi^2_{n_k}(s^2_{m_k} - s^2_{n_k})/s^2_{m_k} \leqq 2\phi^2_{n_k}(s^2_{n_{k+1}} - s^2_{n_k})/s^2_{n_k} \leqq 2b,$$

so that also the second term in (6.16) is bounded. To appraise the last term we note that the derivative of $Q_n(x)$ is of the order of magnitude of λ_n. Thus, using the mean-value theorem we obtain by means of (2.6), (2.4), and (6.17)

$$\phi^2_{m_k} \left| Q_{m_k}(\phi_{m_k}) - Q_{m_k}(\phi_{n_k}) \right| < 2\phi^2_{m_k}\lambda_{m_k}(\phi_{m_k} - \phi_{n_k}) < 1,$$

which proves (6.15) and, accordingly, the divergence of (6.1).

The same argument applies in the case of convergence.

7. **Proof of Theorem** 6. By virtue of Theorem 2 it suffices to prove: *If $\phi(t)$ is a non-decreasing function, $\phi(t) > 2$, then the integral* (2.15) *converges if, and only if,*

$$\sum \frac{\sigma_n^2}{s_n^2} \phi(s_n^2) e^{-(1/2)\phi^2(s_n^2)} \tag{7.1}$$

converges. Since $(1/t)\phi(t)e^{-(1/2)\phi^2(t)}$ is nonincreasing we have

$$\int_{s_k^2}^{\infty} \frac{\phi(t)}{t} e^{-(1/2)\phi^2(t)} dt = \sum_{n=k+1}^{\infty} \int_{s_{n-1}^2}^{s_n^2} \frac{\phi(t)}{t} e^{-(1/2)\phi^2(t)} dt$$

$$\geqq \sum_{n=k+1}^{\infty} \frac{s_n^2 - s_{n-1}^2}{s_n^2} \phi(s_n^2) e^{-(1/2)\phi^2(s_n^2)},$$

so that the convergence of (2.15) implies the convergence of (7.1). Similarly, if (2.15) diverges so does

$$\sum \frac{\sigma_n^2}{s_{n-1}^2} \phi(s_{n-1}^2) e^{-(1/2)\phi^2(s_{n-1}^2)}. \tag{7.2}$$

We have to show that the divergence of (7.2) implies that of (7.1). First let us observe that

$$s_n^2 = s_{n-1}^2 + \sigma_n^2 < s_{n-1}^2 + (1/10^4)s_n^2,$$

or

$$s_{n-1}^2 > (1/2)s_n^2. \tag{7.3}$$

Furthermore, let ν_k denote those integers for which

$$\phi(s_{\nu_k}^2) > \phi(s_{\nu_k - 1}^2) + 1/\phi(s_{\nu_k - 1}^2). \tag{7.4}$$

Since $\nu_{k-1} \leqq \nu_k - 1$ this implies that $\phi^2(s_{\nu_k}^2) > \phi^2(s_0) + 2k$, which, together with (7.3), shows that

$$\sum_k \frac{\sigma_{\nu_k}^2}{s_{\nu_k - 1}^2} \phi(s_{\nu_k - 1}^2) e^{-(1/2)\phi^2(s_{\nu_k - 1}^2)} \in \mathcal{C}.$$

Thus, by assumption, the partial sum in (7.2) extended over $n \neq \nu_1, \nu_2, \cdots$, diverges; however, for these n

$$(\sigma_n^2/s_{n-1}^2)\phi(s_{n-1}^2)e^{-(1/2)\phi^2(s_{n-1}^2)} < 2e^3(\sigma_n^2/s_n^2)\phi(s_n^2)e^{-(1/2)\phi^2(s_n^2)},$$

which completes the proof.

8. **Proof of Theorem** 11. Writing, for abbreviation,

$$\eta_n = 2(\log\log s_n)^{1/2} \tag{8.1}$$

we have from (2.4) that

$$(\eta_n^2/2)\{1 + Q_n(\eta_n)\} \geqq (3/8)\eta_n^2$$

so that

$$\sum \frac{\sigma_n^2}{s_n^2}\eta_n e^{-(1/2)\eta_n^2(1+Q_n(\eta_n))} < 2\sum \frac{\sigma_n^2}{s_n^2}\frac{(\log\log s_n)^{1/2}}{\log^{3/2} s_n}. \tag{8.2}$$

As $\sum\sigma_n^2$ diverges, it is seen that the right-hand member of (8.2) is convergent([17]). Now, by (2.25), the sequence $\{\eta_n\}$ satisfies the condition of Theorem 1. Hence $\{\eta_n\} \in \mathcal{U}$.

The convergence of (8.2) implies that the two series

$$\sum \frac{\sigma_n^2}{s_n^2}\phi_n \exp\{-(1/2)\phi_n^2[1 + Q_n(\phi_n)]\} \tag{8.3}$$

and

$$\sum \frac{\sigma_n^2}{s_n^2}\psi_n \exp\{-(1/2)\psi_n^2[1 + Q_n(\psi_n)]\} \tag{8.4}$$

either both converge or both diverge: as a matter of fact, those terms in (8.3) which are different from the corresponding terms in (8.4) form a subseries of the left-hand member in (8.2).

On the other hand, if $\{\psi_n\} \in \mathcal{U}$, then also $\{\phi_n\} \in \mathcal{U}$, since $\psi_n \leqq \phi_n$. Again, $\psi_n \leqq \eta_n$ and $\{\eta_n\} \in \mathcal{U}$; accordingly, if the inequality $S_n > s_n\psi_n$ is satisfied for infinitely many n, this will be the case for such n for which $\psi_n = \phi_n$. Hence, if $\{\psi_n\} \in \mathcal{L}$, also $\{\phi_n\} \in \mathcal{L}$.

Appendix

The purpose of this appendix is to show how the previous results can be extended, in many cases, to sequences of unbounded variables X_n. The main application is, of course, to the case where all X_n have the same distribution

([17]) Cf. Knopp, loc. cit.

function (cf. the corollary at the end). As a typical example we consider only Theorem 2. The method applies also to the remaining theorems.

From now on the variables $\{X_n\}$ are no longer necessarily bounded; however, we still assume that the second moments (1.3) are finite and that (1.4) holds.

THEOREM. *Suppose that for some constant* $M>0$

$$\sum_k \frac{1}{s_k^2}(\log\log s_k)^2 \int_{|x|>M\cdot s_k/(\log\log s_k)^{1/2}} x^2 dV_k(x) \in \mathcal{C}. \tag{1}$$

Then $\{\phi_n\}\in U(\mathcal{L})$ *if, and only if,*

$$\sum \frac{\sigma_n^2}{s_n^2}\phi_n e^{-(1/2)\phi_n^2} \in \mathcal{C}(\mathfrak{D}). \tag{2}$$

(Needless to say the criterion of Theorem 6 also holds.)

Proof. It follows from the last two paragraphs of §8 that it is sufficient to prove the theorem for sequences $\{\phi_n\}$ such that

$$\phi_n \leqq 2(\log\log s_n)^{1/2}. \tag{3}$$

Put

$$a_n = \int_{|x|\leqq M\cdot s_n/(\log\log s_n)^{1/2}} x dV_n(x)$$

$$X_n^* = \begin{cases} X_n - a_n & \text{if } |X_n| \leqq M\, s_n/(\log\log s_n)^{1/2}, \\ a_n & \text{if } |X_n| > M\, s_n/(\log\log s_n)^{1/2}, \end{cases} \tag{4}$$

$$X_n^{**} = \begin{cases} 0 & \text{if } |X_n| \leqq M\, s_n/(\log\log s_n)^{1/2}, \\ X_n & \text{if } |X_n| > M\, s_n/(\log\log s_n)^{1/2}. \end{cases}$$

Obviously

$$X_n = X_n^* + X_n^{**} + a_n. \tag{5}$$

Now

$$\sum_n \Pr(X_n^{**} \neq 0) = \sum_n \int_{|x|>M\cdot s_n/(\log\log s_n)^{1/2}} dV_n(x)$$

$$\leqq \frac{1}{M^2}\sum_n \frac{(\log\log s_n)^2}{s_n^2} \int_{|x|>M\cdot s_n/(\log\log s_n)^{1/2}} x^2 dV_n(x) < \infty,$$

so that at almost all points

$$\left|\sum_n X_n^{**}\right| = O(1) = o(s_n/\phi_n). \tag{6}$$

Next, by (4),

$$|a_n| \leqq Ms_n/(\log\log s_n)^{3/2}, \tag{7}$$

and therefore by definition

$$|X_n^*| \leqq 2Ms_n/(\log\log s_n)^{3/2}. \tag{8}$$

The variance of X_k^* is given by

$$\begin{aligned}\sigma_k^{*2} &= \int_{|x|\leqq M\cdot s_k/(\log\log s_k)^{3/2}} (x-a_k)^2 dV_k(x) + a_k^2 \int_{|x|>M\cdot s_k/(\log\log s_k)^{3/2}} dV_k(x) \\ &= \int_{|x|\leqq M\cdot s_k/(\log\log s_k)^{3/2}} x^2 dV_k(x) - a_k^2 \qquad (9)\\ &= \sigma_k^2 - \int_{|x|>M\cdot s_k/(\log\log s_k)^{3/2}} x^2 dV_k(x) - a_k^2.\end{aligned}$$

Now it follows from (1.4) and (1) that

$$\begin{aligned}\frac{(\log\log s_n)^{3/2}}{s_n}\sum_{k=1}^n |a_k| &\leqq \sum_{k=1}^n \frac{(\log\log s_k)^{3/2}}{s_k}\int_{|x|>M\cdot s_k/(\log\log s_k)^{3/2}} |x|\, dV_k(x) \\ &\leqq \frac{1}{M}\sum_{k=1}^n \frac{(\log\log s_k)^3}{s_k^2}\int_{|x|>M\cdot s_k/(\log\log s_k)^{3/2}} x^2 dV_k(x) \\ &= O(1).\end{aligned}$$

Accordingly

$$\sum_{k=1}^n |a_k| = O\left(\frac{s_n}{(\log\log s_n)^{3/2}}\right) \tag{10}$$

and, by (7),

$$\sum_{k=1}^n |a_k^2| = O\left(\frac{s_n^2}{(\log\log s_n)^3}\right). \tag{11}$$

Similarly

$$\begin{aligned}&\sum_{k=1}^n \int_{|x|>M\cdot s_k/(\log\log s_k)^{3/2}} x^2 dV_k(x) \\ &\qquad \leqq \frac{s_n^2}{(\log\log s_n)^3}\sum_{k=1}^n \frac{(\log\log s_k)^3}{s_k^2}\int_{|x|>M\cdot s_k/(\log\log s_k)^{3/2}} x^2 dV_k(x) \\ &\qquad = O\left(\frac{s_n^2}{(\log\log s_n)^3}\right).\end{aligned}$$

Combining this with (3), (9), and (11) we have

$$(12)\qquad s_n^{*2} = \sum_{k=1}^{n} \sigma_k^{*2} = s_n^2 + O\left(\frac{s_n^2}{(\log\log s_n)^\delta}\right) = s_n^2 + O\left(\frac{s_n^2}{\phi_n^\delta}\right).$$

This implies that the two sequences $\{\phi_n\}$ and $\{(s_n/s_n^*)\phi_n\}$ belong to the same class with respect to $\{X_n^*\}$. Now, if $\{\phi_n\}\in\mathcal{L}$, at almost all points the inequality

$$(13)\qquad S_n^* = \sum_{k=1}^{n} X_k^* > s_n^*\{(s_n/s_n^*)\phi_n + C(s_n^*/s_n\phi_n)\}$$

will be satisfied for any $C>0$ for infinitely many n so that $\{\phi_n\}\in\mathcal{L}$ also with respect to $\{X_n\}$. A similar remark applies if $\{\phi_n\}\in\mathcal{U}$. In view of (5), (9) the last inequality implies that also $S_n>s_n\phi_n$ for infinitely many n. This finishes the proof.

COROLLARY. *The criteria of Theorems 2 and 6 are valid if there exist two constants $\delta>0$ and A such that*

$$(14)\qquad \int_{-\infty}^{+\infty} x^2 |\log|x||^{1+\delta} dV_k(x) < A\sigma_k^2.$$

In particular, this is the case if $V_k(x)\equiv V(x)$ and

$$(15)\qquad \int_{-\infty}^{+\infty} x^2 |\log|x||^{1+\delta} dV(x) < \infty.$$

In fact, if (14) holds, the series (1) is (at least for sufficiently large k) majorated by

$$C\sum_k \frac{\sigma_k^2}{s_k^2 \log^{1+\delta/2} s_k},$$

which series obviously converges.

REFERENCES

1. F. P. Cantelli, *Considerazioni sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del sig. Paul Lévy*. Giornale dell'Istituto Italiano degli Attuari vol. 4 (1933) pp. 327–350.

2. P. Erdös, *On the law of the iterated logarithm*, Ann. of Math. (2) vol 43 (1942) pp. 419–436.

3. W. Feller, *Generalization of a probability limit theorem of Cramér*, Trans. Amer. Math. Soc. vol. 54 (1943) pp. 361–372.

4. G. H. Hardy and J. E. Littlewood, *Some problems of diophantine approximation*, Acta Math. vol. 37 (1914) pp. 155–239. (Cf., in particular, p. 185.)

5. P. Hartman and A. Wintner, *On the law of the iterated logarithm*, Amer. J. Math. vol. 63 (1941) pp. 169–176.

6. F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1913.

7. A. Khintchine, *Ueber dyadische Brüche*, Math. Zeit. vol. 18 (1923) pp. 109–116.

8. ———, *Ueber einen Satz der Wahrscheinlichkeitsrechnung*, Fund. Math. vol. 6 (1924) pp. 9–20.

9. ———, *Ueber das Gesetz der grossen Zahlen*, Math. Ann. vol. 96 (1926) pp. 156–168.

10. A. Kolmogoroff, *Ueber das Gesetz des iterierten Logarithmus*, Math. Ann. vol. 101 (1929) pp. 126–135.

11. P. Lévy, *Sur un théorème de M. Khintchine*, Bull. Sci. Math. (2) vol. 55 (1931) pp. 145–160.

12. ———, *Théorie de l'addition des variables aléatoires*, Paris, Gautier-Villars, 1937.

13. ———, *Sur les séries dont les termes sont des variables éventuelles indépendantes*, Studia Mathematica vol. 3 (1931) pp. 117–155.

14. J. Marcinkiewicz, *Quelques théorèmes de la théorie des probabilités*, Travaux de la Société des Sciences et des Lettres de Wilno, Classe des Sciences Mathématiques et Naturelles vol. 13 (1939) pp. 1–13.

15. J. Marcinkiewicz and A. Zygmund, *Remarque sur la loi du logarithme itéré*, Fund. Math. vol. 29 (1937) pp. 215–222.

16. H. Steinhaus, *Les probabilités dénombrables et leur rapport à la théorie de la mesure*, Fund. Math. vol. 4 (1922) pp. 286–310.

17. J. Ville, *Étude critique de la notion de collectif*, Paris, Gautier-Villars.

BROWN UNIVERSITY,
PROVIDENCE, R. I.

REGULARITY OF LABEL-SEQUENCES UNDER CONFIGURATION TRANSFORMATIONS

BY

T. N. E. GREVILLE(1)

1. **Introduction.** It is the purpose of this paper to develop a set of transformations to be applied to a label-sequence, or a set of such sequences, yielding a single, new sequence, and to show that certain essential properties of the sequences are almost always preserved under such transformations. This result, like other less general theorems previously given by other writers, provides a new kind of justification for the use of the classical rules for the combination of probabilities. These transformations are defined by setting up certain correspondences between configurations of elements in the original sequences and single elements in the sequence resulting from the transformation. The set of transformations considered will be shown to include as special cases the four fundamental operations in terms of which the rules for the combination of probabilities have been formulated analytically by von Mises(2), as well as the operations used by Copeland, Reichenbach, and Popper in defining classes of canonical sequences.

The concept of the configuration transformation was first suggested by a consideration of the problem of "sets of games." A limited application of the principle to certain simple cases was made in an earlier paper in which it was stated, on the authority of von Mises himself, that his methods were not applicable to the formulation of such problems(3). He has since pointed out(4), however, that the general "set-of-games" problem can be handled by means of a suitable combination of the operations of selection, mixing, and combination. A more general class of problems can be dealt with through the transformations of this paper; and, in any case, it is convenient to work with a single type of operation which includes all the others as special cases. Moreover, it simplifies the treatment of complicated problems to express them in

Presented to the Society, September 8, 1942, and also April 7, 1939 under the title *Invariance of the admisssibility of variates under certain general types of transformations*; received by the editors January 29, 1943, and, in revised form, March 31, 1943.

(1) A dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in the University of Michigan. The writer wishes to express his appreciation to Professor A. H. Copeland for his valuable assistance in the preparation of this paper, and to Professor R. von Mises for permitting him to read the manuscript of an unpublished paper [3]. Numbers in brackets refer to literature cited at end of paper.

(2) von Mises [1, pp. 75–94].

(3) Greville [1, p. 410]; von Mises [1, pp. 108–109].

(4) von Mises [2, pp. 195–197; 3].

terms of a single operator rather than a large (and frequently infinite) number.

My results as to the invariance of certain properties of sequences under configuration transformations include as special cases most of the theorems previously proved on the existence of the collective. For the methods used in arriving at these results, I am much indebted to Kolmogoroff, Feller, and Doob([5]).

2. **The configuration transformation.** Let $[\pi_1, \pi_2, \cdots]$ be a set (finite or denumerable) of arbitrary sets, each containing more than one element, which will be called *label-spaces*. Let $x=[x_1, x_2, \cdots]$ be a set of infinite label-sequences, $x_i=x_i^{(1)}, x_i^{(2)}, \cdots$, each sequence x_i being composed of elements of the set π_i.

A transformation on the set x will be defined by means of a *configuration function*([6]) $f(c)$, in which c denotes a finite configuration of elements of the product-space $\pi=\pi_1\times\pi_2\times\cdots$. It is assumed that $f(c)$ is defined for some, but not necessarily all, finite configurations of elements of π. The "values" of $f(c)$ may be any arbitrary elements.

From the sequence of configurations: $c_1, c_2, \cdots$, in which c_r is the configuration formed by the first r elements of all the sequences x_i, select the subsequence $c_{r_1}, c_{r_2}, \cdots$ consisting of those configurations c_r for which $f(c_r)$ is defined. Finally, form the sequence $y=T(x)=f(c_{r_1}), f(c_{r_2}), \cdots$. This defines the configuration transformation T. The label-space ρ, consisting of all possible "values" of $f(c)$, will be assumed to contain more than one element.

As an example of a configuration transformation, consider two players simultaneously throwing dice, who agree that A is to win a bet if he throws a one before B throws a six. Otherwise, B wins; but if both throw their numbers on the same toss, it is not counted. Here, π_1 and π_2 each consist of the integers 1 to 6, and π consists of all the possible pairs of such integers, while ρ consists of the two labels A and B. The configurations for which the function $f(c)$ is defined are those finite sequences of pairs of integers 1 to 6 terminating with a pair whose first member is 1 *or* whose second member is 6, but not both. $f(c)$ has the value A in the former case, B in the latter.

The four fundamental operations of von Mises, place selection, partition, mixing, and combination, are all particular types of configuration transformations. The first three are all operations on a single sequence. A place selection can be defined([7]) by means of a configuration function $g(c)$ defined for all finite configurations, and restricted to the values 1 and 0. A sequence x' is selected from the sequence x by adopting the rule that any element $x^{(Q}$ is

([5]) Kolmogoroff [1], Feller [1], Doob [1].

([6]) Configuration functions have previously been used by Wald [1], Doob [1], Feller [1], and Ville [1], as a means of defining position selections.

([7]) Wald [1, pp. 38–39]; Doob [1, pp. 364–365]; Feller [1, p. 89]; Ville [1, pp. 41–42]. See also von Mises [1, p. 75].

selected if and only if $g(x^{(1)}, x^{(2)}, \cdots, x^{(i-1)}) = 1$. This operation can be regarded as a configuration transformation by merely including the element selected in the configuration, considering the label-spaces π and ρ as identical.

The operation of mixing([8]) consists in grouping together certain sets of labels and agreeing to represent all the members of a given group or set by a single label. The operation of partition consists([9]) in selecting from a sequence x a subsequence composed of all those elements belonging to some subset γ of π. The operation of combination([10]) consists in combining a set (finite or denumerable) of sequences $[x_1, x_2, \cdots]$ into a single sequence whose elements are points in a multi-dimensional space having as coordinates the corresponding elements of the individual sequences. All these operations are clearly particular cases of the configuration transformation.

3. **A measure theory.** Every set of label-sequences $x = [x_1, x_2, \cdots]$ may be regarded as a point of the infinite product-space $\Pi = \pi \times \pi \times \cdots$. In order to facilitate the study of the behavior of sequences under configuration transformations, a measure theory on Π will be developed. It will be assumed that there exists a field Φ, consisting of subsets of π and including π itself, and a non-negative, additive set function $p(\gamma)$, defined on Φ, such that $p(\pi) = 1$. It will be assumed further that $p(\gamma)$ is absolutely continuous: that is, if there is a monotone sequence $\gamma_1 \supset \gamma_2 \supset \cdots$ approaching the null set as limit, then

$$\lim_{k \to \infty} p(\gamma_k) = 0.$$

According to a well known extension theorem of Banach([11]), a set function $p(\gamma)$ having the above properties can be extended into an absolutely additive set function defined on a Borel field Φ'([12]). The function so obtained will be called the Borel extension([12]) of the function $p(\gamma)$.

We shall follow Kolmogoroff in calling the set $g = \lambda_1 \times \lambda_2 \times \cdots \times \lambda_n \times \pi \times \pi \times \cdots$, where each λ_i is an element of π, a *cylinder-set*([13]) on the space Π; and we define $P(g) = p(\lambda_1)p(\lambda_2) \cdots p(\lambda_n)$. Lomnicki and Ulam([14]) have shown that a set function thus defined can be extended into an absolutely additive (and therefore absolutely continuous) measure function defined on a Borel field F consisting of subsets of Π. Obviously, $P(\Pi) = 1$.

If the sequences $x_1, x_2, \cdots$ are independent, it is possible to begin by

([8]) von Mises [1, p. 76].

([9]) von Mises [1, p. 86].

([10]) von Mises [1, p. 94].

([11]) See Kolmogoroff [1, pp. 15–16], Feller [1, p. 90].

([12]) A Borel field is one having the property that every set which is the sum of a denumerable infinity of sets of the field belongs to the field. See Kolmogoroff [1, pp. 15–16].

([13]) Kolmogoroff [1, p. 25] uses the term *cylinder-set* in a somewhat broader sense. We shall therefore designate as a *simple cylinder-set* a set of the simple product type used here. See also Feller [1, p. 90].

([14]) Lomnicki and Ulam [1, p. 252]. Kolmogoroff [1, p. 27] has also proved this theorem for the case of a Euclidean space. See also Doob [1].

assuming the existence on each space π_i of a set function $p_i(\gamma_i)$, defined on a field Φ_i, having the required properties. A unique measure function $p(\gamma)$ on a field Φ is then obtained by the same process used in developing the function $P(g)$. $p(\gamma)$ automatically possesses the necessary properties.

In order for this measure on Π to be of use in the study of configuration transformations, it is necessary to limit consideration to configuration functions $f(c)$ which are measurable. Consider a finite product-set $s=\lambda_1\times\lambda_2\times\cdots\times\lambda_n$([15]) on the n-dimensional product-space $\rho\times\rho\times\cdots\times\rho$. Let E_s denote the point set on Π consisting of sets of sequences $x=[x_1, x_2, \cdots]$ such that the configuration formed by the first n elements of the corresponding sequence y belongs to s. Further, let E_{sm} denote the subset of E_s consisting of those points x for which the configuration c associated with the element $y^{(n)}$ (through the relation $f(c)=y^{(n)}$) contains m elements. If there exists a field Ψ, consisting of subsets of ρ including ρ itself, and such that, for every such set s formed from elements of Ψ and for every m, the corresponding set E_{sm} belongs to the field F, the configuration function $f(c)$ is said to be *measurable* on Ψ, and the transformation T is called a *measurable transformation*. In the case of a selection, this definition can be shown to be equivalent to that given by Feller([16]).

4. **The regularity theorem.** In studying the behavior of sequences under configuration transformations, we are interested in the frequency with which certain sets of elements appear in the transformed sequence. We shall develop a measure function on the transformed label-space ρ, based on the structure of the transformation T, which will indicate what frequencies are to be expected for these sets on an a priori basis. The measure theory already developed on Π will make it possible to determine whether our expectations are satisfactorily met. In general we shall consider this to be the case if the set of untransformed sequence-sets not producing the proper frequencies in the transformed sequence is of measure zero on Π. In order to secure a suitable measure function on the transformed space, it is necessary to impose certain further (although as shown later, not too severe) limitations on the transformations which may be considered.

A measurable configuration transformation T will be called *endometric* if there exists an additive set function $q(\delta)$, defined for every set δ of Ψ, such that $q(\rho)\leqq 1$, and having the property that, for every finite product-set $s=\lambda_1\times\lambda_2\times\cdots\times\lambda_n$ formed from sets of Ψ,

$$P(E_s)\leqq q(\lambda_1)q(\lambda_2)\cdots q(\lambda_n). \tag{1}$$

The function $q(\delta)$ will be called the *expected frequency* of the set δ. Let

([15]) We shall always use the ordinary multiplication sign to indicate a multi-dimensional product-set and the dot to indicate the set-product or common part of a group of sets.

([16]) Feller [1, p. 90].

$\theta_n(\delta, y)$ denote the number of elements, out of the first n elements of y, which belong to δ.

THEOREM 1. *If T is an endometric configuration transformation, and δ is a set of the field Ψ, then, for almost every point x of Π, either the sequence $y=T(x)$ terminates after a finite number of elements, or else*[17]

$$\lim_{N\to\infty} N^{-1}\theta_N(\delta, y) = q(\delta). \tag{2}$$

Let $M_{n,k}$ denote the set of points of Π for which the sequence $y=T(x)$ contains at least n elements and exactly k of the first n elements belong to δ. Let η denote the set $\rho-\delta$. Then, since T is endometric, $q(\rho)=q(\delta)+q(\eta)\leqq 1$. Hence $q(\eta)\leqq 1-q(\delta)$. It follows from (1) that

$$P(M_{n,k}) \leqq C_{n,k}[q(\delta)]^k[1-q(\delta)]^{n-k}.$$

By a well known property of the binomial expansion[18], the measure of the set

$$\sum_{n=N}^{\infty} \sum_{|k/n-q(\delta)|>\epsilon} M_{n,k}$$

approaches zero as N tends to infinity, for every positive ϵ. Therefore the point set on Π for which the upper or the lower limit of $N^{-1}\theta_n(\delta, y)$ differs from $q(\delta)$ by more than ϵ has the measure zero.

This theorem leaves something to be desired, since it implies that a transformed sequence terminating after a finite number of elements is quite as satisfactory as an infinite sequence having the expected frequency for the set δ. By imposing certain additional requirements on T, this defect can be eliminated. A measurable transformation T will be called *complete* if, for almost every point x of Π, the sequence $y=T(x)$ contains an infinite number of elements. A transformation will be called *holometric* if it is both endometric and complete. A point x on Π such that the sequence $y=T(x)$ does not terminate, and such that (2) holds for *every* set δ of Ψ will be called *a regular point under* T. A set function $f(a)$ defined on a field A will be called *denumerably ordered* if there exists a denumerable set of sets b_k $(k=1, 2, \cdots)$, all contained in A, such that for every set a of A[19],

$$f(a) = \text{G.L.B.}_{b_k\supset a} f(b_k) = \text{L.U.B.}_{b_k\subset a} f(b_k). \tag{3}$$

The following theorem follows immediately from Theorem 1 and from the above definitions:

---

(17) This theorem was proved by Feller [1, pp. 89–93] for the case in which T is a selection.

(18) Borel [1, p. 4] Feller [1, p. 93].

(19) This condition was used by Feller [1, p. 89]. G. L. B. denotes "greatest lower bound"; L. U. B., "least upper bound."

THEOREM 2. *If a configuration transformation T is holometric with respect to a function $q(\delta)$ which is denumerably ordered on a field Ψ, then almost every point of Π is regular under T.*

5. **Iteration of configuration transformations.** It follows from the definition of a configuration transformation that if $y=T(x)$ and $z=T'(y)$, T' being one-dimensional, then there exists a configuration transformation T'' such that $z=T''(x)$. Therefore, any finite sequence of configuration transformations in which all but the first are one-dimensional can be replaced by a single configuration transformation.

In order to discuss the behavior of y under T', it is necessary to define a satisfactory measure function on the label-space ρ. The requirement that T be endometric is not quite adequate, as it does not impose sufficient limitation on $q(\delta)$ to permit its use as a measure function. For this reason, another special type of configuration transformation will now be defined. An endometric transformation will be called *isometric* if, for every product-set $s=\lambda_1\times\lambda_2\times\cdots\times\lambda_n$ formed from sets of Ψ,

$$P(E_s) = q(\lambda_1)q(\lambda_2)\cdots q(\lambda_n). \tag{4}$$

It will be shown later that every holometric transformation is isometric (see Theorem 4). Therefore, in the case of a holometric transformation, the function $q(\delta)$ meets all the requirements for a satisfactory measure function, and can therefore be extended, as indicated in §4, into an absolutely additive set function defined on a Borel field Ψ. It is then possible to define a measure function $Q(d)$ on a field G consisting of subsets of the infinite product space $R=\rho\times\rho\times\cdots$.

Returning to the discussion of iterated transformations, it will now be shown that if T is holometric and T' is measurable, then T'' is measurable. Let σ denote the label-space associated with the sequence z, and let $S=\mu_1\times\mu_2\times\cdots\times\mu_n$ denote a finite product-set on the n-dimensional product-space $\sigma\times\sigma\times\cdots\times\sigma$. Further, let H_S denote the set of points y on R such that the configuration formed by the first n elements of z belongs to S; and let H_{S^m} denote the subset of H_S consisting of those points y for which the configuration c associated with the element $z^{(n)}$ contains m elements. Since T' is measurable, there exists a field Ξ composed of subsets of σ, such that for every m and for every product-set S formed from elements of Ξ, H_{S^m} is measurable on R. Now, there corresponds to every sequence y the set of sequence-sets x such that $T(x)=y$. Similarly, there corresponds to every set of points y on R a set of points x on Π. If K denotes any point set on R and L the corresponding point set on Π, we shall presently show that L is measurable provided K is measurable, and moreover $P(L)=Q(K)$. But if the set H_{S^m} be taken as the set K, the set L is the set E''_{S^m} associated with the configuration S under the transformation T''. This not only shows that T'' is measurable if T' is measurable, but also that if T' is endometric, complete,

holometric, or isometric, then T'' has the same property.

To prove the above statement regarding corresponding point sets on the spaces R and Π, first consider the case in which K is a simple cylinder-set $\lambda_1 \times \lambda_2 \times \cdots \times \lambda_n \times \rho \times \rho \times \cdots$, each λ_i being an element of Ψ'. In this case, it follows from (4) that $P(L) = Q(K)$, since T, being holometric, is therefore isometric (see Theorem 4). Now, the measure function $Q(d)$ on the entire field G was secured by extension from its definition for simple cylinder-sets. Moreover, Lomnicki and Ulam have shown that this extension is unique([20]). As the two measure functions $P(g)$ and $Q(d)$ have the same laws of combination, the same equality must hold for all measurable sets.

A sequence of configuration transformations applied in succession to a sequence-set x, so that we have $T_n\{\cdots T_2[T_1(x)]\} = z$, will be called a *chain*. It follows, of course, that all the transformations of the chain, with the possible exception of T_1, are one-dimensional. The following theorem about chains is an immediate consequence of the above remarks.

THEOREM 3. *A chain of configuration transformations in which the final transformation is measurable, and all the others holometric, is equivalent to a single, measurable configuration transformation. If the final transformation of the chain is endometric, complete, holometric, or isometric, then the equivalent single transformation has the same property.*

6. **Classification of configuration transformations.** Consider the division of any configuration c for which $f(c)$ is defined into two segments or subconfigurations which shall have the following properties. The first segment shall either be null, or else shall constitute a configuration for which $f(c)$ is defined. The second segment shall not be vacuous but otherwise shall contain as few elements as possible. It is clear that such a division is always possible, and is unique. The two segments so obtained will be called the *initial segment* and the *final segment* of c. If a configuration function $f(c)$ is a function of the *final segment only* of c, the corresponding transformation T is called *symmetric*. The transformation associated with the "set-of-games" problem is evidently symmetric, as are all mixings, partitions, and combinations.

On the other hand, place selections are not, in general, symmetric. An example is a type of selection employed by various writers([21]) in defining special classes of canonical sequences, an operation which selects every element in a sequence which is preceded by a specified configuration. Here two configurations on which selections are based may overlap, in which case the transformation is not symmetric. It is easily verified that the class of symmetric transformations possesses the iterative property previously mentioned.

The following are the principal theorems regarding the classification of configuration transformations:

([20]) Lomnicki and Ulam [1, pp. 250–251].

([21]) Reichenbach [1, p. 291], Popper [1, p. 106], Ville [1, p. 70].

THEOREM 4. *Every holometric configuration transformation is isometric.*

Consider the finite product-set $s=\lambda_1\times\lambda_2\times\cdots\times\lambda_n$ in which each set λ_i is an element of the field Ψ associated with the transformation T; and let $\mu_i=\rho-\lambda_i$. Then if $s'=(\lambda_1+\mu_1)\times(\lambda_2+\mu_2)\times\cdots\times(\lambda_n+\mu_n)$, the corresponding point set $E_{s'}$ on Π has unit measure, since the only points x not contained in this set are those for which $T(x)$ terminates with fewer than n elements. Since T is complete, the set of these points is of measure zero. Now, the product s' can be expanded algebraically into a sum of 2^n mutually exclusive product sets, $s_1, s_2, \cdots, s_{2^n}$. In order to have the sum of their measures unity, the measure of each of the corresponding sets, $E_{s_1}, E_{s_2}, \cdots, E_{s_{2^n}}$ (of which one is the set E_s) must have its maximum value according to (1).

THEOREM 5. *Every measurable symmetric transformation is isometric.*

Condition (4) can be satisfied, for the case $n=1$, by defining $q(\delta)=P(E_\delta)$. Assume that (4) holds for all $(n-1)$-dimensional product-sets, and consider the set $s=\lambda_1\times\lambda_2\times\cdots\times\lambda_n$. Let s' denote the set $\lambda_1\times\lambda_2\times\cdots\times\lambda_{n-1}$. Now, the set $E_{s'_m}$ can, by definition, be expressed in the form $K_m\times\pi\times\pi\times\cdots$, where K_m is a subset of the m-dimensional product-space $\pi\times\pi\times\cdots\times\pi$. Let E_δ denote the set of points x such that the first element of y is an element of δ, δ being any subset of ρ. Then, $E_{s'_m}\cdot E_s=K_m\times E_{\lambda_n}$, where the latter expression represents the set of all points x (regarded as sequences of points of the space π) which can be formed by prefixing to any sequence x belonging to the set E_{λ_n} the configuration formed by the m elements constituting any point of K_m. It follows from the manner in which the function $P(g)$ was constructed that $P(K_m\times E_{\lambda_n})=P(E_{s'_m})P(E_{\lambda_n})$. But $E_s=\sum_{m-1}^{\infty}E_{s'_m}\cdot E_s$; and since, by definition, the sets $E_{s'_m}$ for different values of m are mutually exclusive, $P(E_s)=\sum_{m-1}^{\infty}P(E_{s'_m})P(E_{\lambda_n})=P(E_{s'})q(\lambda_n)$. This completes the induction.

THEOREM 6. *Every measurable selection is endometric*[22].

Since, in the case of a selection, the label-spaces π and ρ are identical, take $q(\gamma)=p(\gamma)$. First, consider the case $n=1$, where $s=\lambda_1$. If x belongs to E_{sm}, it follows from the definition of a selection that the first $m-1$ elements of x determine whether or not the mth element will be selected, and that the only role of the mth element, so far as membership in the set E_s is concerned, is in determining whether the element selected belongs to the set λ_1. Therefore, the set E_{sm} can be expressed in the form $K_{m-1}\times\pi\times\pi\times\cdots$, where K_{m-1} is a subset of the $(m-1)$-dimensional space $\pi\times\pi\times\cdots\times\pi$. Then, $E_{sm}=K_{m-1}\times L$, where L denotes the infinite product-set $\lambda_1\times\pi\times\pi\times\cdots$: that is, the set of all sequences x having as the first element a point of λ_1.

(22) Feller [1, pp. 92–93] gives a similar proof for the case in which each component λ_i of s is either a particular set γ or the complement $\pi-\gamma$.

Then, $P(E_{sm}) = P(E_{\tau m})p(\lambda_1)$. Therefore,

$$P(E_s) = \sum_{m=1}^{\infty} P(E_{sm}) = \sum_{m=1}^{\infty} P(E_{\tau m})p(\lambda_1) = P(E_\tau)p(\lambda_1) \leqq p(\lambda_1).$$

This proves the theorem for the case $n=1$.

Now suppose the theorem to be true for all $(n-1)$-dimensional product-sets and consider a product-set s of n dimensions. Let s' denote the product-set obtained by replacing the final element λ_n of s by π, and let s'' denote the $(n-1)$-dimensional product-set obtained by deleting this element. Then, by reasoning similar to that employed in the case $n=1$, $P(E_{sm}) = P(E_{s'm})p(\lambda_n)$. Therefore,

$$P(E_s) = \sum_{m=1}^{\infty} P(E_{sm}) = \sum_{m=1}^{\infty} P(E_{s'm})p(\lambda_n) = P(E_{s'})p(\lambda_n).$$

Since $E_{s'}$ is clearly a subset of $E_{s''}$, $P(E_{s'}) \leqq P(E_{s''})$, whence $P(E_s) \leqq P(E_{s''})q(\lambda_n)$. But, by hypothesis, (1) holds for the product-set s''. This completes the induction.

The properties of the four fundamental operations of von Mises are of particular interest. First considering measurability, every combination is measurable if there exists a measure function on π satisfying the requirements previously laid down, for the field Ψ can then be chosen as identical with Φ. The partition is measurable, provided the set γ on which the selection is based (see §2) belongs to the field Φ, for in this case the set γ together with all its subsets which are also elements of Φ constitutes a field Ψ which is identical with a subfield of Φ. The measure function $q(\delta)$ is defined by

$$q(\delta) = p(\delta)/p(\gamma).$$

The mixing is measurable provided there exists on the space ρ a field Ψ into which some subfield of Φ is transformed by the mixing. No general statement can be made regarding the measurability of place selections.

The mixing, partition, and combination, being all symmetric transformations, are, by Theorem 5, isometric when measurable. Every measurable mixing or combination is evidently complete. Finally, all measurable partitions are complete (unless $p(\gamma)=0$), since it is easily shown that the set of points x on Π containing only a finite number of elements of γ is of measure zero.

7. **Illustration.** An illustration of the application of configuration transformations is to be found in the definition given by Copeland(22) for an *admissible variate*. Let the space π consist of all real numbers, and let ρ consist of the two labels 0 and 1. Let $r_1, r_2, \cdots, r_k, n$ be a set of integers such that $0<r_1<r_2< \cdots <r_k \leqq n$, and let $I_1, I_2, \cdots, I_k$ be a set of real intervals,

(22) Copeland [1, pp. 543–547].

open on the left and closed on the right. The configuration function $f(c)$ will be defined for those and only those configurations in which the number of elements is an integral multiple of n. Its value is determined by the last n elements of the configuration, and is 1 if and only if the r_ith of these last n elements falls in the interval I_i for each i ($i=1, 2, \cdots, k$). We shall call any transformation T thus defined a *Copeland transformation*. If $F(t)$ is any monotonic function such that $F(-\infty)=0$ and $F(\infty)=1$, a label-sequence x is said to be an admissible variate with respect to the function $F(t)$ provided the relative frequency of the label 1 in the corresponding sequence $y=T(x)$ tends to the limit $\prod_{i=1}^{k}\int_{I_i}dF(t+0)$, for every possible choice of the integers $r_1, r_2, \cdots, r_k, n$ and of the intervals I_i.

Given any function $F(t)$ satisfying the above conditions, Copeland shows that there exists an admissible variate with respect to $F(t)$. We shall show that almost every variate x is admissible with respect to any such function $F(t)$. Let the field Φ consist of all real intervals I open on the left and closed on the right, and all finite sums of such intervals; and for each such interval I, let $p(I)=\int_I dF(t+0)$. Consider the Copeland transformation T resulting from a particular choice of the integers $r_1, r_2, \cdots, r_k, n$ and the intervals I_i. This transformation is clearly measurable, and being symmetric and complete, is therefore holometric. By Theorem 2, almost every sequence x is regular under any denumerable set of Copeland transformations. As $F(t)$ is a monotonic function, it has at most a denumerable set of discontinuities. Let the set G consist of all those intervals I whose end points are either rational points or points of discontinuity of $F(t)$. Now let the set D consist of all possible Copeland transformations based only on intervals belonging to G. D is then a denumerable set, and almost every sequence x is regular under every transformation of D. But such a sequence is regular under all Copeland transformations, since, for any interval I, it is possible to select a sequence of intervals $I_1, I_2, \cdots$, all belonging to G, such that

$$\lim_{m\to\infty}\int_{I_m} dF(t+0)=\int_I dF(t+0).$$

8. **General remarks.** Although the theorems of Wald([24]) are more general than those of this paper in that they apply to *all* selections, no measure theory is developed for appraising sets of sequences, and transformations other than selections are not considered. Feller([25]) avoids the assumption of absolute continuity and the limitation to measurable selections by constructing a modified (and somewhat arbitrary) measure function in terms of which all configuration functions are measurable. It is not known whether this principle can be extended to transformations other than selections.

([24]) Wald [1].
([25]) Feller [1].

There are certain interesting non-endometric transformations, which nevertheless possess the regularity property of Theorem 2. No necessary and sufficient condition for this property is at present known. Nor is it known whether the general (non-symmetric) configuration transformation can be replaced by a set of operations of the von Mises types.

References

E. Borel

1. *Traité du calcul des probabilités et de ses applications*, vol. 2, no. 1, chap. 1, Paris, 1926.

A. H. Copeland

1. *A matrix theory of measurement*, Math. Zeit. vol. 37 (1933) pp. 542–555.

J. L. Doob

1. *Note on probability*, Ann. of Math. vol. 37 (1936) pp. 363–367.

W. Feller

1. *Über die Existens von sogenannten Kollektiven*, Fund. Math. vol. 32 (1939) pp. 87–96.

T. N. E. Greville

1. *Invariance of the admissibility of numbers under certain general types of transformations*, Trans. Amer. Math. Soc. vol. 46 (1939) pp. 410–425.

A. Kolmogoroff

1. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse Mathematische, vol. 2, no. 3, Berlin, 1933.

Z. Lomnicki and S. Ulam

1. *Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités. I. Variables indépendantes*, Fund. Math. vol. 23 (1934) pp. 237–278.

R. von Mises

1. *Vorlesungen aus dem Gebiete der angewandte Mathematik*, vol. 1, *Wahrscheinlichkeit*, Leipzig, 1931.
2. *On the foundations of probability and statistics*, Ann. Math. Statist. vol. 12 (1941) pp. 191–205.
3. *On the probabilities in a set of games and the foundation of probability theory*, unpublished manuscript.

K. Popper

1. *Logik der Forschung*, Vienna, 1935.

H. Reichenbach

1. *Les fondements logiques du calcul des probabilités*, Annales de l'Institut Henri Poincaré vol. 7 (1937) pp. 267–348.

J. Ville

1. *Étude critique de la notion de collectif*, Paris, 1939.

A. Wald

1. *Die Widerspruchsfreiheit des Kollektivsbegriffes der Wahrscheinlichkeitsrechnung*, Ergebnisse eines mathematischen Kolloquiums vol. 8 (1935–1936) pp. 38–72.

Bureau of the Census,
Washington, D. C.

THE THEORY OF ANALYTIC FUNCTIONS IN NORMED ABELIAN VECTOR RINGS

BY

EDGAR R. LORCH

Introduction. The problem of determining the broadest domain over which the classic results of analytic function theory are valid has attracted attention continuously since the time of Hamilton's essays in that direction with quaternions(1). The contributions to this subject converge from many directions; matrix theory, the theory of hypercomplex numbers, the theory of integral equations, and others, provide the frame for the investigations(2). It is established in the present paper that classical function theory can be extended to normed abelian vector rings. This result was known for finite dimensional rings(3). The infinite dimensional case requires for its elaboration certain basic information concerning the structure of bounded linear transformations. A surprising feature of the present development is that the methods as well as the results of the classic theory can be carried over almost without change(4).

In the preliminary section are derived the elementary properties of the rings here considered. Then the Cauchy theory is elaborated. The theory of ideals is examined insofar as it bears strongly on present questions. Use is made frequently of the notion of maximal ideal; the existence and principal properties of these were established by Gelfand(5). Illustrative of the results here obtained are Theorems 6 and 9. The exponential and logarithmic func-

Presented to the Society, February 28, 1942, and April 24, 1943; received by the editors November 23, 1942.

(1) In his *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert* (Julius Springer, Berlin, 1927), Felix Klein writes: "Das letzte Ziel (of quaternion theory) war—und ist—eine quaternionistische Funktionentheorie, von der man ganz neue, gewaltige umfassende Aufschlüsse für die ganze Mathematik erhoffte. Zur Förderung dieser zwar nicht deutlichen aber gläubig angenommenen Ziele wurde gar 1895 ein 'Weltbund zur Förderung der Quaternionen' gegründet." See vol. 1, p. 189.

(2) A large though necessarily incomplete list of references is given by J. A. Ward, *A theory of analytic functions in linear associative algebras*, Duke Math. J. vol. 7 (1940) pp. 232–248.

(3) G. Scheffers studied functions analytic over commutative and associative algebras with a finite basis—*Verallgemeinerung der Grundlagen der gewöhnlichen komplexen Funktionen*, Leipziger Berichte vol. 45 (1893) pp. 838–848; vol. 46 (1894) pp. 120–134. Scheffers' results were extended by P. W. Ketchum, *Analytic functions of hypercomplex variables*, Trans. Amer. Math. Soc. vol. 30 (1928) pp. 641–667.

(4) A very characteristic change in the present development is that the equality for complex numbers $|a \cdot b| = |a| \cdot |b|$ is replaced by the inequality in rings $|a \cdot b| \leq |a| \cdot |b|$. It can readily be seen that certain classic proofs of function theory (convergence of integrals and series) are not affected by this change.

(5) I. Gelfand, *Normierte Ringe*, Rec. Math. (Mat. Sbornik) N.S. vol. 9 (1941) pp. 1–24.

tions are discussed. The importance of considerations of reducibility here may be judged from the fact that exp (z) is simply periodic if and only if the ring is irreducible. The problem of embedding an element in a continuous multiplicative group is treated completely in Theorem 11. Finally, separation in a ring by a rectifiable curve is discussed in the last section: separation and irreducibility are interrelated concepts. The present paper is based on an analysis of spectra of linear transformations which the author has recently published([6]). At certain points theorems are stated without proof as similar results appear elsewhere in the literature. In this respect the work of Gelfand is particularly to be cited. There is interest in noting that certain of his results (for example the theory of the radical) may be obtained equally from our work by the simple device of obtaining a representation of the given abstract ring by means of a ring of transformations (Theorem 1).

Vector rings. A collection $\mathfrak{R}$ of elements $a, b, z, \cdots$ is called a *normed abelian vector ring* if([7]):

(a) $\mathfrak{R}$ is a normed linear complex vector space.

(b) $\mathfrak{R}$ is a commutative ring with unit e.

(c) The norm function defined over $\mathfrak{R}$ satisfies the inequality

$$|a \cdot b| \leqq |a| \cdot |b|. \tag{1}$$

Elementary properties of these rings will be derived rapidly. The inequality $|ab-ac| \leqq |a| \cdot |b-c|$ shows that multiplication is continuous in the topology defined in $\mathfrak{R}$ by the norm. If a^{-1} represents the reciprocal (in transformation theory, the inverse) of a, the equation

$$(a - z)^{-1} = a^{-1} + a^{-2}z + a^{-3}z^2 + \cdots,$$

valid for $|z| < |a^{-1}|^{-1}$, shows that the elements in $\mathfrak{R}$ possessing an inverse form an open set. It also shows that inversion is a continuous (indeed, analytic) procedure. If a^{-1} exists, a is called *regular*; in the contrary case, a is called *singular*. The regular elements form a topological group G relative to multiplication. The set G, being open, is the point set sum of disjoint maximal open connected sets G_α, $G=\sum_\alpha G_\alpha$([8]). That component containing the unit e will be denoted by G_1 and will be called the principal component of G. If $a \in G_1$, $b \in G_1$, then $ab \in G_1$ and $a^{-1} \in G_1$, that is, G_1 is a group. The set F of

([6]) *The spectrum of linear transformations*, Trans. Amer. Math. Soc. vol. 52 (1942) pp. 238–248. This paper will be referred to as S. In preparing S, the author unfortunately relied too heavily on the literature of the past decade to furnish him with a list of references. For that reason it was only after the present manuscript was submitted that he noted the brief but pointed results of F. Riesz in initiating this theory. These are to be found in his book, *Les systèmes d'équations linéaires*, Paris, 1913, pp. 117–121. I hope Professor Riesz will accept my sincere apologies.

([7]) This definition appears in a paper by M. Nagumo, *Einige analytische Untersuchungen in linearen metrischen Ringen*, Jap. J. Math. vol. 13 (1936) pp. 61–80.

([8]) An example of a ring in which the multiplicative group G is not connected is the ring $\mathfrak{R}_3$ of the example discussed after Theorem 12.

singular elements is closed. If $\mathfrak{J}$ is an ideal, $\mathfrak{J} \neq \mathfrak{R}$, then $\mathfrak{J} \subset F$. If $\mathfrak{J} = F$, $\mathfrak{J}$ is the radical of $\mathfrak{R}$ (see Theorem 6).

If a is fixed in $\mathfrak{R}$, z arbitrary, the mapping

$$T_a z = az$$

establishes a correspondence $a \to T_a$ between $\mathfrak{R}$ and a class of bounded linear transformations defined over $\mathfrak{R}$; this correspondence is called the *regular representation* of $\mathfrak{R}$.

THEOREM 1. *The mapping $a \to T_a$ of the regular representation of $\mathfrak{R}$ is an algebraic and topological isomorphism.*

This theorem enables one to apply the theory of linear transformations to $\mathfrak{R}$. Thus one may say, "the complex number λ is in the spectrum of a" instead of the more cumbersome "λ is in the spectrum of T_a"(9).

The ring $\mathfrak{R}$ is *reducible* if $\mathfrak{R}$ is the disjoint sum of two rings each distinct from $\mathfrak{R}$, $\mathfrak{R} = \mathfrak{R}_1 + \mathfrak{R}_2$. More precisely, this means that if $a \in \mathfrak{R}$, there exists a unique pair $a_1 \in \mathfrak{R}_1$, $a_2 \in \mathfrak{R}_2$ such that $a = a_1 + a_2$, $a_1 a_2 = 0$. Reducibility implies and is implied by the existence of an idempotent j, $j^2 = j$, with $j \neq 0$, $j \neq e$. The rings $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are given by $\mathfrak{R}_1 = j\mathfrak{R}$, $\mathfrak{R}_2 = (e-j)\mathfrak{R}$ where $e-j$ is also an idempotent. A fundamental characterization of irreducible rings is given in the following theorem.

THEOREM 2. *The ring $\mathfrak{R}$ is irreducible if and only if the spectrum of every element in $\mathfrak{R}$ is a connected set.*

If $\mathfrak{R}$ is reducible, and if j is an idempotent in $\mathfrak{R}$, $j \neq e$, $j \neq 0$, the spectrum of j consists of the two points $\lambda = 0$, $\lambda = 1$, a disconnected set. Now let a be any element in $\mathfrak{R}$ whose spectrum is not connected. Let C be a simple closed rectifiable curve in the resolvent set of a such that points of the spectrum of a lie both interior and exterior to C. The integral

$$j = \frac{1}{2\pi i} \int_C \frac{d\zeta}{\zeta e - a} \tag{2}$$

evaluated over C in the counterclockwise sense has the property $j^2 = j$(10). By S, Theorem 8c, d, p. 244, $j \neq 0$, $j \neq e$, hence $\mathfrak{R}$ is reducible.

Clearly, if $\mathfrak{R}$ is finite dimensional, it is the direct sum of irreducible rings. This is not the case for infinite dimensional rings. Irreducible rings are fundamental in function theory. For example, the ring of continuous complex

(9) Throughout the paper, complex numbers will be denoted by Greek letters $\alpha, \gamma, \zeta, \lambda, \cdots$. It is worth while to compare certain expressions of abstract algebra and the corresponding expressions in transformation theory. Thus "a is a null-divisor" is equivalent to "$\lambda = 0$ is in the point spectrum of a."

(10) See S, Theorem 6, p. 242. That the spectrum can be separated by a curve in case it is not connected is discussed under *Spectral components*, p. 245.

functions defined on a finite closed real interval, or the ring of functions analytic and bounded over a region of the complex plane are irreducible([11]).

It is important to know what (complex) rings $\mathfrak{R}$ may be fields. The answer is provided by Theorem 3. It is also interesting to determine what rings have the property $|a \cdot b| = |a| \cdot |b|$. This is settled by Theorem 4.

THEOREM 3. *If $\mathfrak{R}$ is a field, it is the field of complex numbers*([12]).

THEOREM 4. *If in $\mathfrak{R}$ the norm function satisfies $|ab| = |a| \cdot |b|$ then $\mathfrak{R}$ is the field of complex numbers.*

Using the fact that $|e| = 1$, one proves that if $b \in \mathfrak{R}$ is regular, the spectrum of b lies on the circle $|\zeta| = |b|$. Since this is also true of $b - \mu e$ where $|\mu|$ is small, the spectrum of b consists of one point. Applying S, Theorems 8d and 7, it may be shown that b is a scalar.

Analytic functions. Although most proofs will be omitted in this section, it will be valuable to set down salient steps in sufficient detail so that the character of the development is clearly apparent. The purpose of this is to show that not merely the results but also the methods and even the appearance of the classic theory may be transferred to $\mathfrak{R}$.

A variable element in $\mathfrak{R}$ will be denoted by z. A function defined over a region in $\mathfrak{R}$ and having values in $\mathfrak{R}$ will be denoted by $f(z)$. Continuity of $f(z)$ is defined in evident fashion. The topology used throughout is that of the norm, that is, the uniform topology. A function $z = z(t)$ where $z(t)$ is a continuous function defined on a real interval $0 \leqq t \leqq 1$ defines a curve. The curve is rectifiable if the sums $\sum_{i=1}^{n} |z(t_i) - z(t_{i-1})|$ are bounded; here n is arbitrary and $0 = t_0 \leqq t_1 \leqq \cdots \leqq t_n = 1$. It may be shown that if $f(z)$ is continuous in a region H and if C is any rectifiable curve, closed or not, in H, the Riemann integral $\int_C f(z)dz$ exists and has the usual properties.

DEFINITION. *A function $f(z)$ is said to have a derivative $f'(z_0)$ at $z = z_0$ if for any $\epsilon > 0$ a $\delta > 0$ can be found such that for all h in $\mathfrak{R}$ with $|h| < \delta$,*

$$|f(z_0 + h) - f(z_0) - hf'(z_0)| \leqq |h| \epsilon. \tag{3}$$

If $f(z)$ has a derivative throughout a neighborhood of $z = z_0$, $f(z)$ is said to be analytic at $z = z_0$.

([11]) Let $\mathfrak{R}$ be the ring of complex valued continuous functions $f(t)$, $0 \leqq t \leqq 1$, with $|f| = \max |f(t)|$. The intimate connection of the concept of irreducibility with some fundamental theorems of function theory will be illustrated: the spectrum of $f(t)$ consists of the set of values assumed by $f(t)$. If $\mathfrak{R}$ is assumed irreducible, this is a closed connected set. If $f(t)$ assumes real values only, the set is a real closed interval. Thus $f(t)$ assumes its maximum (or minimum) on $0 \leqq t \leqq 1$. Also, $f(t)$ assumes every value between these extremes.

If $\mathfrak{Q}$ is the ring of functions $f(z)$ analytic and bounded on $|z| < 1$ with $|f| = \text{l.u.b.} |f(z)|$, $\mathfrak{Q}$ is a normed abelian vector ring. Since $|z| < 1$ is a connected set and $f(z)$ is continuous the spectrum of $f(z)$ is connected and $\mathfrak{Q}$ is irreducible.

([12]) Theorems 3 and 4 have been given by S. Mazur, C.R. Acad. Sci. Paris vol. 207 (1938) p. 1025.

The Cauchy theorem may be established for functions analytic in a convex region (or in a region which is the sum of a finite number of convex regions). As in the classic case it is first established for triangles. If a and b are in $\mathfrak{R}$, the line segment joining a and b is given by $ta+(1-t)b$, $0 \leqq t \leqq 1$. The interior of the triangle determined by a, b, c consists of the points $\lambda a+\mu b+\nu c$, $\lambda+\mu+\nu=1$, $\lambda>0$, $\mu>0$, $\nu>0$. The proof now proceeds in standard fashion. Incidentally, (3) is required in the form there given and not in the quotient form $|\Delta w/\Delta z-f'(z_0)| \leqq \epsilon$ normally employed to define the derivative. From triangles one proceeds to arbitrary polygons; in a convex region this is trivially simple. Finally the proof is given for an arbitrary rectifiable curve.

The Cauchy formula for $f(z)$ cannot be established without specifying carefully the curve C around which the integration is performed. In some sense the point z must be interior to C. As an illustration of the phenomena which may arise, the integral (2) does not always have the expected value e or 0. However, our development needs only the fact that, for certain special curves C, the classic integral yields $f(z)$. If C is a simple curve lying in the z-plane, that is, in the plane of points $z+\lambda e$ where λ is complex and arbitrary, and if z is interior to C in this plane, then the Cauchy formula holds. It is valid not only for that z but for all points of a neighborhood of z.

Finally one establishes the Taylor expansion for $f(z)$ in the usual manner. It may be seen that in this development analytic functions could be defined equivalently by (1) the uniqueness of the derivative; (2) the uniqueness of the integral (Morera's Theorem); and (3) the possibility of a power series expansion. The concept of conformality is meaningless in $\mathfrak{R}$ except in the case that $\mathfrak{R}$ is the complex number field. For conformality may be defined as follows: $\mathfrak{R}$ is a conformal ring if $|az| = K_a|z|$ where K_a is some constant. Considering $\mathfrak{R}$ a ring of transformations (or equivalently, considering that $|e| = 1$) and setting $z=e$ yields $|az| = |a| \cdot |z|$ for all a, z in $\mathfrak{R}$. By Theorem 4, $\mathfrak{R}$ is the field of complex numbers.

A series of powers of $z-a$ (which converges for some regular $z-a$) is called an element. An analytic function in the large is defined as the totality of elements which are continuations of a given one. It is curious to note that one formula may define more than one analytic function. For example, if G has more than one component, $1/z$ defines an analytic function in each. One may establish Liouville's Theorem for functions analytic and bounded everywhere. From this the fundamental theorem of algebra is obtained in the following form: Given $f(z)=c_nz^n+ \cdots +c_0$, $n>0$, c_n regular, there exists at least one element a in $\mathfrak{R}$ such that $f(a)$ is singular. The existence of a zero is not asserted. If c_n is singular, $f(a)$ may be regular for all a.

It is seen that in this development partial differential equations of the Cauchy-Riemann type play no role. In the infinite dimensional case these equations would be of a most complicated type (being dependent on bases)—

if they could be set up at all. In the finite case, the development has been carried through by means of systems of partial differential equations(13). It would seem that this circumstance contributes to a lack of transparence in the given treatment.

Ideal theory in $\mathfrak{R}$. In what follows, it will be at times useful, at others necessary, to employ ideal-theoretic considerations in $\mathfrak{R}$. The concept of maximal ideal is particularly important. A maximal ideal is one which while distinct from $\mathfrak{R}$ is not contained in any other ideal. Although an ideal is not necessarily a closed set, the word "ideal" will be used subsequently as an abbreviation for "closed ideal." The basic results on ideals have been given by Gelfand and can be summarized as follows(14):

THEOREM 5. *If $\mathfrak{I}$ is any ideal in $\mathfrak{R}$, the ring $\mathfrak{R}/\mathfrak{I}$ is a normed abelian vector ring. An ideal $\mathfrak{K}$ can always be found which contains $\mathfrak{I}$, $\mathfrak{K} \supset \mathfrak{I}$, and which is maximal. The ideal $\mathfrak{I}$ is maximal if and only if $\mathfrak{R}/\mathfrak{I}$ is the field of complex numbers.*

It is of importance to note that many properties of $\mathfrak{R}$ remain invariant under the operation of residuation. An example will be given: if $f(z)$ is a function of z analytic at $z=c$, and if $\mathfrak{I}$ is any ideal in $\mathfrak{R}$, then $\bar{f}(\bar{z})$ is a function of $\bar{z}$ in $\overline{\mathfrak{R}}=\mathfrak{R}/\mathfrak{I}$ analytic at $\bar{z}=\bar{c}$. For from the series

$$f(z) = a_0 + a_1(z-c) + a_2(z-c)^2 + \cdots$$

one obtains

$$\bar{f}(\bar{z}) = \bar{a}_0 + \bar{a}_1(\bar{z}-\bar{c}) + \bar{a}_2(\bar{z}-\bar{c})^2 + \cdots .$$

The latter series converges for some regular $\bar{z}-\bar{c}$ since $|\bar{a}_n(\bar{z}-\bar{c})^n| \leqq |a_n(z-c)^n|$, hence represents an analytic function.

The following results characterize the structure of $\mathfrak{R}$ in terms of the spectra of the elements of $\mathfrak{R}$. It has been seen in Theorem 2 how the reducibility of $\mathfrak{R}$ can be characterized in this fashion. An element a is *nilpotent* if for some positive integer n, $a^n=0$. An element a is called a *quasi-nilpotent* if for all complex μ, $\lim_{n\to\infty}(\mu a)^n=0$.

THEOREM 6. *An element a is a quasi-nilpotent if and only if the spectrum of a consists of the one point $\lambda=0$. The quasi-nilpotent elements constitute an ideal $\mathfrak{K}$ called the radical of $\mathfrak{R}$. The radical is the intersection of all maximal ideals. If $\overline{\mathfrak{R}}=\mathfrak{R}/\mathfrak{K}$ and $\bar{a}\in\overline{\mathfrak{R}}$ is the image of $a\in\mathfrak{R}$, a and $\bar{a}$ have identical spectra. Thus the ring $\mathfrak{R}/\mathfrak{K}$ is without radical.*

If the spectrum of a consists of the point $\lambda=0$, then by S, Theorem 7, $(\mu a)^n\to 0$ and a is a quasi-nilpotent. Also if a is such that $(\mu a)^n\to 0$ the only

(13) See footnote 3.

(14) Gelfand, loc. cit. pp. 6, 8, 9.

singularity of a is $\lambda=0$. The remainder of the proof follows the ideas of Gelfand.

THEOREM 7. *If $\mathfrak{J}$ is an ideal contained in the radical $\mathfrak{R}$ of $\mathfrak{R}$, $\mathfrak{R}/\mathfrak{J}$ is irreducible if and only if $\mathfrak{R}$ is irreducible.*

THEOREM 8. *If $f(\lambda)=\alpha_0+\alpha_1\lambda+\alpha_2\lambda^2+\cdots$ is a power series with scalar coefficients which converges absolutely for some $\lambda\neq 0$ and a is any element in $\mathfrak{R}$ whose spectrum lies entirely within the circle in the complex plane $|\zeta|=|\lambda|$, the series $f(a)$ converges absolutely. The spectrum of $f(a)$ is the transform under $f(\lambda)$ of the spectrum of a.*

Since the spectrum of $\lambda^{-1}a$ lies within the unit circle, by S, Theorem 7, $|(\lambda^{-1}a)^n|=|(\lambda^{-1}a)^n e|\to 0$. Thus for n sufficiently large, $|a^n|<|\lambda^n|$ and $f(a)$ converges absolutely.

Now suppose λ is in the spectrum of a. Let $\mathfrak{J}$ be any ideal such that $a\equiv\lambda e(\mathfrak{J})$. Write $p_n(\lambda)=\alpha_0+\cdots+\alpha_n\lambda^n$, $n=0, 1, \cdots$. Then $p_n(a)\equiv p_n(\lambda)e(\mathfrak{J})$. Letting n increase, one has $f(a)\equiv f(\lambda)e(\mathfrak{J})$ or $f(\lambda)$ is in the spectrum of $f(a)$. Next let μ be in the spectrum of $f(a)$; let $\mathfrak{J}$ be a maximal ideal such that $f(a)\equiv\mu e(\mathfrak{J})$. If $a\equiv\lambda e(\mathfrak{J})$, repetition of the previous argument yields $f(\lambda)=\mu$.

If in particular $\sum_{n=0}^{\infty}|\alpha_n|<\infty$ and $|a|=1$, then once more the spectrum of $f(a)$ is the transform under $f(\lambda)$ of that of a. An application of this result will be given.

Let $f(z)=\sum_{n=0}^{\infty}\alpha_n z^n$ be such that $\sum_{n=0}^{\infty}|\alpha_n|<\infty$. Then if $f(z)\neq 0$ in $|z|\leq 1$, $1/f(z)=\sum_{n=0}^{\infty}\beta_n z^n$ with $\sum_{n=0}^{\infty}|\beta_n|<\infty$.

In proof, let $\mathfrak{R}$ be the ring of all $f=f(z)$ for which $\sum_{n=0}^{\infty}|\alpha_n|<\infty$. Let $|f|=\sum_{n=0}^{\infty}|\alpha_n|$. Then in $\mathfrak{R}$, $|z|=1$ and clearly the spectrum of z consists of all λ with $|\lambda|\leq 1$. Thus by the theorem, the spectrum of f consists of all values assumed by $f(z)$ for $|z|\leq 1$. Since 0 is not one of them, $1/f(z)$ is in $\mathfrak{R}$.

THEOREM 9. *Let $f(\lambda)=\alpha_0+\alpha_1\lambda+\alpha_2\lambda^2+\cdots$ $(\neq 0)$ be a scalar power series convergent for all λ. Let $a\in\mathfrak{R}$ be such that $f(a)=0$. Then there exists a polynomial $p(\lambda)$ $(\neq 0)$ with scalar coefficients such that $p(a)=0$. In particular if $\mathfrak{R}$ is irreducible and without radical, a is a scalar, $a=\mu e$.*

Since the spectrum of a is bounded and the zeros of $f(\lambda)$ are finite in number in any bounded set, the spectrum of a consists of a finite number of isolated points, $\lambda_1, \cdots, \lambda_n$. Let $j_1, \cdots, j_n$ be the idempotents

$$j_k=\frac{1}{2\pi i}\int_{C_k}\frac{d\zeta}{\zeta e-a}$$

where C_k is a circle in the complex plane centered on $\lambda_k e$ and sufficiently small so as to exclude the points $\lambda_s e$, $s\neq k$. Then $e=j_1+\cdots+j_n$ and $j_k\cdot j_s=0$ for $s\neq k$ (by S, Theorems 8d and 5).

Suppose first that $n=1$. Expand $f(\lambda)$ around $\lambda=\lambda_1$, $f(\lambda)=\beta_0+\beta_1(\lambda-\lambda_1)+\cdots$, and $f(a)=\beta_0 e+\beta_1(a-\lambda_1 e)+\cdots$. If $\mathfrak{J}$ is any maximal ideal, it contains $a-\lambda_1 e$ which is in the radical of $\mathfrak{R}$, hence it may be seen by a discussion modulo $\mathfrak{J}$ that an element $\gamma_0 e+\gamma_1(a-\lambda_1 e)+\gamma_2(a-\lambda_1 e)^2+\cdots$ is singular if and only if $\gamma_0=0$. Since $f(a)=0$, $\beta_0=0$. Suppose $\beta_1=0, \cdots, \beta_{s-1}=0, \beta_s\neq 0$. Then $f(a)=(a-\lambda_1 e)^s[\beta_s e+\beta_{s+1}(a-\lambda_1 e)+\cdots]=0$ and $(a-\lambda_1 e)^s=0$. If $\mathfrak{R}$ is irreducible, then certainly $n=1$; if $\mathfrak{R}$ is without radical, then one may take $s=1$.

Now consider n unrestricted. The element $(a-\lambda_k e)j_k$ belongs to the radical of the ring $j_k\mathfrak{R}$ consisting of all elements $j_k a$, $a\in\mathfrak{R}$. For since $j_k^2=j_k$, then by S, Theorem 7, for large μ, hence for all μ, $\mu^n(a-\lambda_k e)^n j_k=[\mu(a-\lambda e)j_k]^n\to 0$ as $n\to\infty$; then Theorem 6 is applied. Now it may be seen that for $k=1, \cdots, n$, the ring $j_k\mathfrak{R}$ is isomorphic with the ring $\overline{\mathfrak{R}}_k=\mathfrak{R}/(e-j_k)$ where the symbol $(e-j_k)$ represents a principal ideal. Hence since $\bar{f}(a)=0$ in $\overline{\mathfrak{R}}_k$, the previous paragraph shows that $(\bar{a}-\lambda_k\bar{e})^{s_k}=0$ in $\overline{\mathfrak{R}}_k$, that is, $(a-\lambda_k e)^{s_k}\equiv 0\ (e-j_k)$, $k=1, \cdots, n$. Thus, writing $p(\lambda)=\prod_{k=1}^{n}(\lambda-\lambda_k)^{s_k}$, $p(a)$ is in the intersection of the ideals $(e-j_k)$, $k=1, \cdots, n$. Since this intersection is the zero ideal, $p(a)=0$.

This theorem shows that an analogue to the Cayley-Hamilton theorem that every matrix satisfies its characteristic equation does not exist in $\mathfrak{R}$. Indeed the equation $f(a)=0$ implies that a is finite dimensional in the sense that the ring it generates is finite dimensional.

The exponential and logarithmic functions. The function $\exp(z)$ is defined by the series $\sum_{n=0}^{\infty} z^n/n!$ which converges for all $z\in\mathfrak{R}$. Its functional equation and the formula for the derivative are readily obtained. For all z, $\exp(z)$ lies in the principal component G_1 of $\mathfrak{R}$. First, the inverse of $\exp(a)$ is $\exp(-a)$; next, by the continuity of $\exp(z)$, if C is a curve joining 0 to a, then the map of C under the transformation $z\to\exp(z)$ joins e to $\exp(a)$. The relationship of the exponential to reducibility is interesting.

THEOREM 10. *The function* $\exp(z)$ *is simply periodic if and only if* $\mathfrak{R}$ *is irreducible. More precisely, the periods of* $\exp(z)$ *are exactly the elements* $2\pi i\sum_{k=1}^{n}\pm j_k$ *where the* j_k *are idempotents, distinct or not.*

Direct substitution in the series for $\exp(z)$ shows that $\exp(2\pi ij)=e$ in case j is any idempotent. Thus if $\exp(z)$ is simply periodic, then $\mathfrak{R}$ has but one idempotent (aside from 0), namely e; hence $\mathfrak{R}$ is irreducible. Now suppose that for a given a in $\mathfrak{R}$, $\exp(2\pi ia)-e=0$. By Theorem 8, the spectrum of a consists of one or more of the points $\lambda=0, \pm 1, \cdots$. If a has more than one point in its spectrum, then $\mathfrak{R}$ is reducible (Theorem 2). Suppose now that a has only one point in its spectrum; by translation one may assume that $\lambda=0$ is that point. Let $b=\exp(\pi ia)$; then $b^2=e$ and $(b+e)/2$ is an idempotent. If $b\neq e, b\neq -e$, $\mathfrak{R}$ is reducible. In the contrary case, since the spectrum of a consists of $\lambda=0$, by Theorem 8 the spectrum of $\exp(\pi ia)+e$

consists of $\lambda=2$, hence $b\neq -e$ but rather $b=e$. Consider now in turn the elements $\exp(\pi ia/2)$, $\exp(\pi ia/4)$, and so on; either $\mathfrak{R}$ is found to be reducible or the analytic function $\exp(2\pi iaz)$ has the value e for $z=e/2^n$, $n=0, 1, \cdots$. The function is identically e and its derivative $2\pi ia$ is 0. This shows that $\mathfrak{R}$ is irreducible if and only if $\exp(z)$ is simply periodic.

In proof of the last statement of the theorem, let $\lambda_1, \cdots, \lambda_n$ be the points of the spectrum of a, λ_k having an integral value. Let $j_1, \cdots, j_n$ be the idempotents defined for a as in Theorem 9. Then $a=aj_1+\cdots+aj_n$. Also $\exp(2\pi iaj_k)=e$ and $aj_k-\lambda_k j_k$ is in the radical of $\overline{\mathfrak{R}}=j_k\mathfrak{R}$, $k=1, \cdots, n$. Applying the preceding paragraph, $aj_k=\lambda_k j_k$ giving $a=\lambda_1 j_1+\cdots+\lambda_n j_n$. Except for notation, this is the statement in the theorem.

The logarithm is defined by

$$\log z=\int_e^z \frac{dz}{z}$$

where the path of integration joining e to z is arbitrary but must consist of regular elements. Thus the logarithm is defined for elements in the principal component only. It may be shown that for any a in $\mathfrak{R}$, the equation $\log z=a$ has solutions. By what precedes and by the obvious properties of the logarithm, $\log e=2\pi i\sum_{k=1}^{n}\pm j_k$ (Theorem 10).

The components of G and continuous groups. The components G_α of the group G of regular elements in $\mathfrak{R}$ have been defined earlier; the principal component G_1 is that containing e. The group G_1 consists precisely of the elements a expressible in the form $a=\exp(b)$, $b\in\mathfrak{R}$. For if $b\in\mathfrak{R}$, then $\exp(b)\in G_1$; if $a\in G_1$, then $\log a$ exists and $a=\exp(\log a)$([15]). For every b there is a b' such that $b'\in G_1$ and $\exp(b)=\exp(b')$. Indeed if n is an integer such that $2\pi n>|b|$, then $b'=2\pi ine+b$ satisfies these conditions. Thus the map of G_1 under the exponential function is G_1 in its entirety.

A set of elements a^s, where s is an exponent ranging over the real or complex numbers or over a more general range, constitutes a continuous additive group if $a^s\cdot a^t=a^{s+t}$ and if the group operation is continuous. In many investigations it is of importance to determine whether a given element in $\mathfrak{R}$ may be embedded in such a group.

THEOREM 11. *An element $a\in\mathfrak{R}$ may be embedded in a continuous additive group, $a^s\cdot a^t=a^{s+t}$, if and only if a is in the principal component G_1 of $\mathfrak{R}$. In that case the whole ring $\mathfrak{R}$ serves as a range of exponents. The groups are all of the form $a^s=\exp(s\log a)$ where $\log a$ is any determination of the logarithm.*

If the elements a^s form a continuous group, then letting s have real values, $0\leqq s\leqq 1$, one may see that a is in G_1. It is known that all such continu-

([15]) This result is given by Nagumo, loc. cit. p. 66, Theorem 3.

ous groups with real exponents are of the form $a^s = \exp(s \log a)$([16]). If now s is allowed to range over $\mathfrak{R}$ an additive group is obtained which is an extension of the given group.

This concept of a general exponent seems to be new. Thus if a is a finite dimensional nonsingular matrix one may consider such expressions as a^a.

THEOREM 12. *If the number of components of the topological group G is greater than one, it is infinite.*

Suppose that t is a regular element in $\mathfrak{R}$ not in the principal component G_1. Then the elements t^n, $n=0, \pm 1, \cdots$, belong to distinct components. For suppose that t^{n_1} and t^{n_2}, $n_1 \geqq n_2$, belong to the same component; then $t^{n_1-n_2}=t^m$ is in G_1. Let t^m be embedded in a continuous group (Theorem 11), and if $m \neq 0$, let $(t^m)^{1/m}=s$. Then if $u=ts^{-1}$, $u^m=e$. The spectrum of u consists at most of the mth roots of unity (Theorem 8). Thus $u \in G_1$, hence $t=su \in G_1$. This implies that $m=0$ and that t^n, $n=0, \pm 1, \cdots$ are in distinct components.

Examples will be given to show how the regularity of an element and also the component in which it belongs are influenced by ring extensions and contractions. If a is regular in $\mathfrak{R}$ and the resolvent set of a is connected, then $a \in G_1$; for a and $a+ne$ with $n > |a|$ lie in the same connected component of G and $a+ne \in G_1$. The converse proposition is false. Let $\mathfrak{R}_1$ be the ring of all complex-valued continuous functions $f(t)$, $0 \leqq t \leqq 1$, with $|f| = \max|f(t)|$. Let $\mathfrak{R}_2$ be the subring of $\mathfrak{R}_1$ generated by $\exp(2\pi it)$ and $\exp(-2\pi it)$. Let $\mathfrak{R}_3$ be the subring of $\mathfrak{R}_2$ generated by $\exp(2\pi it)$ and 1. Then in $\mathfrak{R}_1$, $\exp(2\pi it)$ is regular and since it is of the form $\exp(b)$ where $b \in \mathfrak{R}_1$, it is in the principal component. In $\mathfrak{R}_2$, $\exp(2\pi it)$ is regular but not in the principal component; for if $f(t) \in \mathfrak{R}_2$, $f(0)=f(1)$, hence $2\pi it \notin \mathfrak{R}_2$. In $\mathfrak{R}_3$, $\exp(2\pi it)$ is a singular element. For if one writes $\exp(-2\pi it) - \sum_{s=0}^{n} \alpha_s \exp(2\pi ist) = k(t)$ where n and the α_s are arbitrary, then $|k| = \max|k(t) \exp(2\pi it)| \geqq \left|\int_0^1 k(t) \cdot \exp(2\pi it)dt\right| = 1$. That all points λ with $|\lambda| < 1$ are in the spectrum of $\exp(2\pi it)$ considered as an element of $\mathfrak{R}_3$ is a consequence of the next theorem.

The preceding discussion shows that spectral singularities may at times be removed by ring extensions. If $\mathfrak{R}_2 \subset \mathfrak{R}_1$ and $a \in \mathfrak{R}_2$ and if $a-\lambda e$ is singular in $\mathfrak{R}_2$ but regular in $\mathfrak{R}_1$, λ is called a *removable spectral singularity* of a. In case λ is not removable, the singularity is called a *permanent spectral singularity*.

THEOREM 13. *All points belonging to the frontier of the spectrum of an element a in $\mathfrak{R}$ are permanent spectral singularities of a.*

([16]) The fact that any real continuous group a^s, $-\infty < s < \infty$, is of the form $a^s = \exp(sb)$ is proved by Nagumo, loc. cit. p. 72, Theorem 9. See also D. S. Nathan, *One-parameter groups of transformations in abstract vector spaces*, Duke Math. J. vol. 1 (1935) pp. 518–526.

Since the scalar series $(\zeta-\lambda)^{-1}=\sum_{n=1}^{\infty}\lambda^{n-1}\zeta^{-n}$ converges absolutely for $|\zeta^{-1}|<|\lambda^{-1}|$, then by Theorem 8, $\sum_{n=1}^{\infty}\lambda^{n-1}a^{-n}$ converges absolutely (to $(a-\lambda e)^{-1}$) for any element a whose spectrum lies outside the circle $|\zeta|=|\lambda|$.

Let $\mathfrak{R}_2\subset\mathfrak{R}_1$ and let $a\in\mathfrak{R}_2$. If $a-\mu e$ is regular in $\mathfrak{R}_2$, hence also in $\mathfrak{R}_1$, the convergence or divergence of the series for $([a-\mu e]-\lambda e)^{-1}$ is settled without reference to the ring to which a belongs. If ν is a frontier point of the spectrum of a considered as an element of $\mathfrak{R}_2$, this series cannot converge for any λ with $|\lambda|>|\nu-\mu|$. This implies that $a-\nu e$ is singular in $\mathfrak{R}_1$ hence ν is a permanent spectral singularity(17).

An application of Theorem 13 will be given.

Let the spectrum of $a\in\mathfrak{R}$ consist of real points only. Then the spectrum is invariant under ring extensions and contractions.

Separation in $\mathfrak{R}$ by rectifiable curves. The theory of separation in $\mathfrak{R}$ by means of rectifiable curves resembles strongly that in the complex plane. If $C: z=z(t)$, $0\leqq t\leqq 1$, $z(0)=z(1)$, is a closed continuous rectifiable curve (for definition see above), C is called *simple* if $t_1\not\equiv t_2 \pmod 1$ implies $[z(t_1)-z(t_2)]^{-1}$ exists. If C is such a curve in $\mathfrak{R}$ and $\mathfrak{J}$ is any ideal then $\overline{C}: \bar{z}=\bar{z}(t)$ is such a curve in $\overline{\mathfrak{R}}=\mathfrak{R}/\mathfrak{J}$. For $\overline{C}$ is clearly simple and since the norm of $\bar{z}(t)$ in $\overline{\mathfrak{R}}$ does not exceed the norm of $z(t)$ in $\mathfrak{R}$, $\overline{C}$ is rectifiable. In particular, if $\mathfrak{J}$ is maximal, $\overline{\mathfrak{R}}$ is the field of complex numbers and $\overline{C}$ is a simple closed rectifiable curve in the usual sense.

If $C: z=z(t)$ is as above and $a\in\mathfrak{R}$ is such that $z(t)-a$ is singular for some t, a is called *C-singular*. If $z(t)-a$ is regular for all t, a is called *C-regular*. There exist C-regular elements. For example $a=z(0)+\lambda e$, where $|\lambda|$ is sufficiently large, is C-regular. Since C is a compact set the set of C-singular elements is closed, that of C-regular elements is open. Consider the integral

$$u=\frac{1}{2\pi i}\int_C \frac{dz}{z-a} \tag{4}$$

where a is C-regular. Let b be on C and let $z'=(z-a)/(b-a)$. This substitution leaves the integral (4) unchanged in value and replaces C by a simple curve C' lying in G_1 and containing e. Hence $u=(2\pi i)^{-1}\log e=\sum_{k=1}^{n}\pm j_k$ where the j_k are idempotents. If $\mathfrak{R}$ is irreducible, $u=ne$ where n is an integer (Theorem 10).

Let $\mathfrak{J}$ be a maximal ideal and let $z(t)\equiv\zeta(t)(\mathfrak{J})$. Then under the homomorphism $\mathfrak{R}\to\overline{\mathfrak{R}}=\mathfrak{R}/\mathfrak{J}$, the integral (4) is transformed into a classical integral $\bar{u}$ in the complex plane. Choosing a proper sense along $\overline{C}: \zeta=\zeta(t)$, hence along C, $\bar{u}=0$ or $\bar{u}=1$. Since $\bar{u}=n\bar{e}$, $n=0$ or 1. The C-regular elements

(17) This result is implied by a lemma of G. Silov, *On the extension of maximal ideals*, C.R. (Doklady) Acad. Sci. URSS. N.S. vol. 29 (1940) pp. 83–84. Silov proves that $a-\nu e$ is a generalized divisor of zero, hence ν is a permanent spectral singularity. An element b is called a generalized divisor of zero if there exist elements c_n, $n=1, 2, \cdots$, such that $|c_n|=1$ and $|bc_n|\to 0$ as $n\to\infty$.

for which $u=e$ will be called *C-interior* elements. Those for which $u=0$ will be called *C-exterior*. The C-interior elements form an open set which need not be connected. The same is true of the C-exterior elements. Exterior elements exist: if b is on C, $b+\lambda e$ is exterior to C for large $|\lambda|$. The existence of interior elements will be discussed below. In résumé, one has the following theorem.

THEOREM 14. *If $\mathfrak{R}$ is an irreducible ring, a simple closed rectifiable curve $C: z=z(t)$ partitions by means of the integral* (4) *the elements a of $\mathfrak{R}$ into three classes*:

(1) *The closed set of C-singular elements.*

(2) *The open set of C-exterior elements a for which $u=0$.*

(3) *The open set of C-interior elements a for which $u=e$.*

The existence of interior elements is related to differentiability on C in the following theorem.

THEOREM 15. *If $C: z=z(t)$ is a simple closed rectifiable curve in $\mathfrak{R}$ and if for some t the derivative $z'(t)$ exists and is a regular element, then C has interior points.*

Suppose $z'(t)$ exists and is regular at $t=t_0$; for $t-t_0\neq 0$, write $(z(t)-z(t_0))/(t-t_0)=a+h$ where $a=z'(t_0)$. Suppose $(a-b)^{-1}$ exists for $b\in\mathfrak{R}$, $|b|<r$ where r is some constant. Choose ϵ, $0<\epsilon<r$, and find $v>0$ so that when $0<|t-t_0|<v$, $|h|<\epsilon$. Choose a $\delta>0$ such that for $|z-z(t_0)|<\delta$, $|t-t_0|\leqq v$, $(z-z(t))^{-1}$ exists. Let τ be a real number not equal to 0, $|\tau a|<\delta$. Then $w(t)=z(t)-z(t_0)-i\tau a$ is regular for all t. This is clearly true for $|t-t_0|\leqq v$. For $|t-t_0|<v$, write $w(t)$ in the form $w(t)=(t-t_0-i\tau)[a+(t-t_0)(t-t_0-i\tau)^{-1}h]$ and it is seen here also that $w(t)$ is regular.

If $\mathfrak{J}$ is a maximal ideal in $\mathfrak{R}$, and $z(t)\equiv\zeta(t)(\mathfrak{J})$, it may be seen that $\zeta'(t_0)=\alpha$ exists and $\alpha\neq 0$. The point $\zeta(t_0)+i\tau\alpha$ lies within the curve $\zeta=\zeta(t)$ either for a $\tau>0$ or a $\tau<0$. Thus the integral (4) in $\overline{\mathfrak{R}}=\mathfrak{R}/\mathfrak{J}$ has for this choice of τ the value $\hat{u}=1$. Hence $u=e$ and $z(t_0)+i\tau a$ is interior to C.

BARNARD COLLEGE, COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

TESTS OF STATISTICAL HYPOTHESES CONCERNING SEVERAL PARAMETERS WHEN THE NUMBER OF OBSERVATIONS IS LARGE(1)

BY

ABRAHAM WALD

TABLE OF CONTENTS

1. Introduction 426
2. Assumptions on the density function $f(x, \theta)$ 428
3. The joint limit distribution of $\dot{\theta}_\omega$ 429
4. Reduction of the general problem to the case of a multivariate normal distribution .. 433
5. Tests of simple hypotheses which have uniformly best average power over a family of surfaces 445
6. Tests of simple hypotheses which have best constant power on a family of surfaces ... 450
7. Most stringent tests of simple hypotheses 451
8. Definitions of "best" tests of composite hypotheses 453
9. Tests of linear composite hypotheses which have uniformly best average power over a family of surfaces 455
10. Tests of linear composite hypotheses which have best constant power on a family of surfaces 461
11. Most stringent tests of linear composite hypotheses 461
12. The general composite hypothesis 463
13. Optimum properties of the likelihood ratio test 470
14. Large sample distribution of the likelihood ratio 478
15. Summary 481

1. **Introduction.** In this paper we shall deal with the following general problem: Let $f(x^1, x^2, \cdots, x^r, \theta^1, \cdots, \theta^k)$ be the joint probability density function of the variates (chance variables) $x^1, \cdots, x^r$ involving k unknown parameters $\theta^1, \cdots, \theta^k$. Any set of k values $\theta^1, \cdots, \theta^k$ can be represented by a point θ in the k-dimensional Cartesian space with the coordinates $\theta^1, \cdots, \theta^k$. We shall denote the set of all possible parameter points by Ω. The set Ω is called parameter space. The parameter space Ω may be the whole k-dimensional Cartesian space, or a subset of it. For any subset ω of Ω, we shall denote by H_ω the hypothesis that the parameter point lies in ω. If ω consists of a single point, H_ω is called a simple hypothesis, otherwise H_ω is called a composite hypothesis. In this paper we shall discuss the question of an appropriate test of the hypothesis H_ω based on a large number of independent observations on $x^1, \cdots, x^r$.

For simplicity we shall introduce the following notations: The letter θ or θ_i for any subscript i will denote a point in the parameter space Ω. The letter x

Some of the results contained in this paper were presented to the Society, February 22, 1941 and September 2, 1941; received by the editors March 31, 1943.

(1) Research under a grant-in-aid from the Carnegie Corporation of New York.

will denote the random vector with the components $x^1, \cdots, x^r$, and x_α will denote the random vector with the components $x_\alpha^1, \cdots, x_\alpha^r$ where x_α^i $(i=1, 2, \cdots, r)$ denotes the αth observation on x^i. In general, the components of a vector v in the s-dimensional space will be denoted by $v^1, \cdots, v^s$, that is the components of a vector will always be indicated by superscripts. Throughout this paper all vectors have their initial points at the origin. We denote by E_n a sample point in the rn-dimensional sample space of n independent observations on the random vector x. For any relation R we denote by $P(R|\theta)$ the probability that R holds under the assumption that θ is the true parameter point. A region (subset) of the rn-dimensional sample space will always be denoted by a capital letter with the subscript n. For any region W_n the symbol $P(W_n|\theta)$ will denote the probability that E_n falls within W_n under the assumption that θ is the true parameter point. Throughout this paper the word "region" will be used synonymously with "subset," since in the theory of testing statistical hypotheses it is customary to call the subsets which are used as criterions of rejection, critical regions.

By maximum likelihood estimates $\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k$ of $\theta^1, \cdots, \theta^k$ we mean values of $\theta^1, \cdots, \theta^k$ for which $\prod_{\alpha=1}^n f(x_\alpha^1, \cdots, x_\alpha^r, \theta^1, \cdots, \theta^k)$ becomes a maximum. The subscript n in the symbol $\hat{\theta}_n^i$ will indicate that the maximum likelihood estimate is based on n independent observations on $x', \cdots, x^r$. A region W_n in the rn-dimensional sample space is called a critical region for testing the hypothesis H_ω if we decide to reject H_ω when and only when the observed sample point falls within W_n. For any θ not in ω the value of $P(W_n|\theta)$ is called the power of the critical region W_n with respect to the alternative hypothesis θ. The least upper bound of $P(W_n|\theta)$ with respect to θ, restricting θ to ω, will be called the size of the critical region W_n. A critical region is considered the better, the smaller its size and the greater its power.

In several previous publications(2) the author has considered the case of a single unknown parameter θ and the problem of testing a simple hypothesis $\theta=\theta_0$. It was shown, among other things, that under certain conditions the critical region given by the inequality $|n^{1/2}(\hat{\theta}_n-\theta_0)| \geqq A_n$ has certain optimum properties. Here the symbol A_n denotes some properly chosen constant. In this paper the general case of several unknown parameters is treated and simple as well as composite hypotheses are considered.

By an equality or inequality among vectors we shall mean that the equality or inequality holds for all components. For example, if $\hat{\theta}_n$ denotes the vector with the components $\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k$, where $\hat{\theta}_n^i$ is the maximum likelihood estimate of θ^i, and if A is a real number, then the inequality

$$n^{1/2}(\hat{\theta}_n - \theta) < A$$

(2) *Asymptotically most powerful tests of statistical hypotheses*, Ann. Math. Statist. vol. 12 (1941). *Some examples of asymptotically most powerful tests*, Ann. Math. Statist. ibid.

denotes the set of inequalities

$$n^{1/2}(\hat{\theta}_n^i - \theta^i) < A \qquad (i = 1, \cdots, k).$$

Or, if t is a vector with the components $t^1, \cdots, t^k$, then

$$n^{1/2}(\hat{\theta}_n - \theta) < t$$

denotes the set of inequalities

$$n^{1/2}(\hat{\theta}_n^i - \theta^i) < t^i \qquad (i = 1, \cdots, k).$$

2. **Assumptions on the density function $f(x,\theta)$.** For any function $\psi(x)$ we shall use the symbol $\int_{-\infty}^{+\infty}\psi(x)dx$ as an abbreviation for $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\psi(x)\,dx^1\cdots dx^r$. Denote by $E_\theta\psi(x)$ the expected value of $\psi(x)$ under the assumption that θ is the true parameter point, that is

$$E_\theta\psi(x) = \int_{-\infty}^{+\infty}\psi(x)f(x,\theta)dx.$$

For any x, any positive value δ, and any θ_1 denote by $\psi_{ij}(x, \theta_1, \delta)$ the greatest lower bound, and by $\phi_{ij}(x, \theta_1, \delta)$ the least upper bound of $\partial^2 \log f(x, \theta)/\partial\theta^i\partial\theta^j$ with respect to θ in the θ-interval $|\theta-\theta_1| \leqq \delta$.

Throughout this paper the following assumptions on $f(x, \theta)$ are made:

ASSUMPTION I. *Denote by D_n the set of all sample points E_n for which the maximum likelihood estimate $\hat{\theta}_n = (\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k)$ exists and the second order partial derivatives $\partial^2 f(x_\alpha, \theta)/\partial\theta^i\partial\theta^j$ $(\alpha = 1, \cdots, n; i, j = 1, \cdots, k)$ are continuous functions of θ. It is assumed that*

$$\lim_{n=\infty} P(D_n \mid \theta) = 1 \qquad \textit{uniformly in } \theta.$$

If for a sample point E_n there exist several maximum likelihood estimates, we can select one of them by some given rule. Hence we shall consider $\hat{\theta}_n$ as a single-valued function of E_n defined for all points of D_n.

ASSUMPTION II. *For any positive ϵ*

$$\lim_{n=\infty} P[|\hat{\theta}_n - \theta| < \epsilon \mid \theta] = 1$$

uniformly in θ, where $\hat{\theta}_n$ denotes the vector with the components $\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k$ and $\hat{\theta}_n^i$ is the maximum likelihood estimate of θ^i.

Assumption II is somewhat more than consistency of the maximum likelihood estimate $\hat{\theta}_n$. In fact consistency means only that for any positive ϵ

$$\lim_{n=\infty} P[|\hat{\theta}_n - \theta| < \epsilon \mid \theta] = 1,$$

without requiring that the convergence be uniform in θ. If $\hat{\theta}_n$ satisfies Assump-

tion II we shall say that $\hat{\theta}_n$ is a uniformly consistent estimate of θ. A rigorous proof of the consistency of $\hat{\theta}_n$ (under certain restrictions on $f(x, \theta)$) was given by J. L. Doob(3). The uniform consistency of $\hat{\theta}_n$ together with the uniform consistency of the likelihood ratio test will be proved on the basis of some weak assumptions on $f(x, \theta)$ in a forthcoming paper.

ASSUMPTION III. *The following three conditions are fulfilled*:

(a) *For any sequences* $\{\theta_{1n}\}, \{\theta_{2n}\}$, *and* $\{\delta_n\}$ *for which* $\lim_{n=\infty} \theta_{1n} = \lim_{n=\infty} \theta_{2n} = \theta$ *and* $\lim \delta_n = 0$ *we have*

$$\lim_{n=\infty} E_{\theta_{1n}} \psi_{ij}(x, \theta_{2n}, \delta_n) = \lim_{n=\infty} E_{\theta_{1n}} \phi_{ij}(x, \theta_{2n}, \delta_n) = E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j}$$

uniformly in θ.

(b) *There exists a positive* ϵ *such that the expectations* $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ *and* $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ *are bounded functions of* θ_1, θ_2 *and* δ *in the domain* D_ϵ *defined by the inequalities* $|\theta_1 - \theta_2| \leqq \epsilon$ *and* $|\delta| \leqq \epsilon$.

(c) *The greatest lower bound with respect to* θ *of the absolute value of the determinant of the matrix* $\| -E_\theta(\partial^2 \log f(x, \theta))/\partial\theta^i \partial\theta^j \|$ *is positive.*

ASSUMPTION IV. $\int_{-\infty}^{+\infty} \partial f(x, \theta)/\partial\theta^i \, dx = \int_{-\infty}^{+\infty} \partial^2 f(x, \theta)/\partial\theta^i \partial\theta^j \, dx = 0$.

Assumption IV simply means that we may differentiate with respect to θ under the integral sign. In fact

$$\int_{-\infty}^{+\infty} f(x, \theta) dx = 1$$

identically in θ. Hence

$$\frac{\partial}{\partial\theta^i} \int_{-\infty}^{+\infty} f(x, \theta) dx = \frac{\partial^2}{\partial\theta^i \partial\theta^j} \int_{-\infty}^{+\infty} f(x, \theta) dx = 0.$$

Differentiating under the integral sign, we obtain the relations in Assumption IV.

ASSUMPTION V. *There exists a positive* η *such that*

$$E_\theta \left| \frac{\partial}{\partial\theta^i} \log f(x, \theta) \right|^{2+\eta} \qquad (i = 1, \cdots, k)$$

are bounded functions of θ.

3. **The joint limit distribution of $\hat{\theta}_n$.** Denote $n^{1/2}(\hat{\theta}_n^i - \theta^i)$ by $z_n^i(\theta^i)$ and let $z_n(\theta)$ be the vector with the components $z_n^1(\theta^1), \cdots, z_n^k(\theta^k)$. For any constant vector t denote the probability $P[z_n(\theta) < t \,|\, \theta]$ by $\Phi_n(t \,|\, \theta)$. We shall prove the following proposition.

(3) J. L. Doob, *Probability and statistics*, Trans. Amer. Math. Soc. vol. 36 (1934) pp. 759–775.

PROPOSITION I. *The distribution function* $\Phi_n(t|\theta)$ *converges with* $n \to \infty$, *uniformly in* t *and* θ, *towards the cumulative multivariate normal distribution with zero means and covariance matrix*

$$\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$$

where

$$c_{ij}(\theta) = -E_\theta \partial^2 \log f(x, \theta)/\partial\theta^i \partial\theta^j.$$

Proof. Because of Assumption IV, we have

$$E_\theta \frac{\partial \log f(x, \theta)}{\partial \theta^i} = \int_{-\infty}^{+\infty} \frac{\partial f(x, \theta)}{\partial \theta^i} dx = 0. \tag{1}$$

From

$$\frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} = \frac{1}{f(x, \theta)} \frac{\partial^2 f(x, \theta)}{\partial\theta^i \partial\theta^j} - \frac{1}{f^2}\left[\frac{\partial f}{\partial\theta^i} \frac{\partial f}{\partial\theta^j}\right]$$

we obtain because of Assumption IV

$$E_\theta\left\{\left[\frac{\partial \log f}{\partial\theta^i} \frac{\partial \log f}{\partial\theta^j}\right]\right\} = -E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} = c_{ij}(\theta). \tag{2}$$

From (2) it follows that the matrix $\|c_{ij}(\theta)\|$ is positive definite or semidefinite. Because of condition (c) of Assumption III the matrix $\|c_{ij}(\theta)\|$ must be positive definite. For any point E_n of the set D_n defined in Assumption I consider the Taylor expansion

$$\sum_\alpha \frac{\partial \log f(x_\alpha, \theta_1)}{\partial\theta^i} = \sum_\alpha \frac{\partial \log f(x_\alpha, \theta)}{\partial\theta^i} + \sum_j (\theta_1^j - \theta^j)\left[\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j}\right], \tag{3}$$

where $\bar{\theta}$ lies on the segment connecting the points θ and θ_1. Denote $n^{-1/2}\sum_\alpha \partial \log f(x_\alpha, \theta)/\partial\theta^i$ by $y_n^i(\theta)$ and let $y_n(\theta)$ be the vector with the components $y_n^1(\theta), \cdots, y_n^k(\theta)$. Substituting θ_n for θ_1 in (3), the left-hand side of (3) becomes equal to zero and we obtain

$$y_n^i(\theta) + \sum_j \left\{[n^{1/2}(\theta_n^j - \theta^j)] \frac{1}{n}\left[\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j}\right]\right\} = 0 \tag{4}$$

or

$$y_n^i(\theta) + \sum_j z_n^j(\theta) \frac{1}{n}\left[\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j}\right] = 0. \tag{5}$$

Let ν be an arbitrary positive number and let $Q_n(\theta)$ be the subset of D_n for which the inequality

$$\left| \frac{1}{n} \sum_{\alpha} \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} + c_{ij}(\theta) \right| < \nu \tag{6}$$

holds. We will prove that

$$\lim_{n=\infty} P[Q_n(\theta) \mid \theta] = 1 \tag{7}$$

uniformly in θ.

Let τ_0 be a positive number such that

$$\begin{aligned} &\left| E_\theta \phi_{ij}(x, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \frac{\nu}{2}, \\ &\left| E_\theta \psi_{ij}(x, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \frac{\nu}{2} \end{aligned} \tag{8}$$

for all values of θ. Because of condition (a) of Assumption III, such a τ_0 certainly exists. Denote by $R_n(\theta)$ the subset of D_n consisting of all points E_n for which the inequality

$$|\theta_n - \theta| \leqq \tau_0 \tag{9}$$

holds. Because of Assumption II

$$\lim_{n=\infty} P[R_n(\theta) \mid \theta] = 1 \tag{10}$$

uniformly in θ. Since $\bar{\theta}$ lies in the interval $[\theta_n, \theta]$ we have for all points of $R_n(\theta)$

$$|\bar{\theta}^i - \theta^i| \leqq \tau_0 \qquad (i = 1, \cdots, k). \tag{11}$$

Hence at any point of $R_n(\theta)$ the inequality

$$\sum_{\alpha} \psi_{ij}(x_\alpha, \theta, \tau_0) \leqq \sum_{\alpha} \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} \leqq \sum_{\alpha} \phi_{ij}(x_\alpha, \theta, \tau_0) \tag{12}$$

holds. Let the region $S_n(\theta)$ be defined by the inequality

$$\left| \frac{1}{n} \sum_{\alpha} \phi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \phi_{ij}(x, \theta, \tau_0) \right| < \frac{\nu}{2} \tag{13}$$

and $T_n(\theta)$ be defined by the inequality

$$\left| \frac{1}{n} \sum_{\alpha} \psi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \psi_{ij}(x, \theta, \tau_0) \right| < \frac{\nu}{2}. \tag{14}$$

It follows from (b) of Assumption III and Tshebysheff's inequality that

$$\lim_{n=\infty} P[S_n(\theta) \mid \theta] = \lim_{n=\infty} P[T_n(\theta) \mid \theta] = 1 \tag{15}$$

uniformly in θ. Denote by $U_n(\theta)$ the common part of the regions $R_n(\theta)$, $S_n(\theta)$ and $T_n(\theta)$. In $U_n(\theta)$ we have because of (8), (13) and (14)

$$\begin{aligned} &\left| \frac{1}{n} \sum_{\alpha} \phi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \nu, \\ &\left| \frac{1}{n} \sum_{\alpha} \psi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \nu. \end{aligned} \tag{16}$$

From this we obtain (6) because of (12). That is to say, the inequality (6) is valid everywhere in $U_n(\theta)$. Since $\lim_{n=\infty} P[U_n(\theta) \mid \theta] = 1$ uniformly in θ and $U_n(\theta)$ is a subset of D_n, our statement about $Q_n(\theta)$ is proved.

Since the determinant $|c_{ij}(\theta)|$ has a positive lower bound, we obtain easily from (5) and (6)

$$z_n^i(\theta) = \sum_j y_n^j(\theta)[\sigma_{ij}(\theta) + \nu\epsilon_{ijn}(E_n, \theta, \nu)], \tag{17}$$

where $\epsilon_{ijn}(E_n, \theta, \nu)$ is a bounded function of E_n, θ, and ν, provided that, for each θ, E_n is restricted to points of $Q_n(\theta)$ and $|\nu|$ is less than a certain positive number ν_0.

Let $\bar{z}_n(\theta)$ be defined as follows: $\bar{z}_n(\theta) = z_n(\theta)$ at any point of $Q_n(\theta)$, and

$$\bar{z}_n^i(\theta) = \sum_j y_n^j(\theta)\sigma_{ij}(\theta)$$

at any point outside $Q_n(\theta)$. It follows from (7) that

$$\lim_{n=\infty} \{P[\bar{z}_n(\theta) < t \mid \theta] - P[z_n(\theta) < t \mid \theta]\} = 0 \tag{18}$$

uniformly in t and θ.

Denote $\sum_j y_n^j(\theta)\sigma_{ij}(\theta)$ by $\dot{z}_n^i(\theta)$ and let $\dot{z}_n(\theta)$ be the vector with the components $\dot{z}_n^1(\theta), \cdots, \dot{z}_n^k(\theta)$. From (1), (2), Assumption V and the general limit theorems it follows that $P[y_n(\theta) < t \mid \theta]$ converges with $n \to \infty$, uniformly in t and θ, towards the k-variate cumulative normal distribution with zero means and covariance matrix $\|c_{ij}(\theta)\|$. From this we easily obtain that $P[\dot{z}_n(\theta) < t \mid \theta]$ converges with $n \to \infty$, uniformly in t and θ, towards the cumulative joint normal distribution with zero means and covariance matrix $\|\sigma_{ij}(\theta)\|$. Since ν can be chosen arbitrarily small, we obviously have because of (17)

$$\lim_{n=\infty} \{P[\dot{z}_n(\theta) < t \mid \theta] - P[\bar{z}_n(\theta) < t \mid \theta]\} = 0 \tag{19}$$

uniformly in t and θ. Proposition I follows from (18) and (19).

4. Reduction of the general problem to the case of a multivariate normal distribution. In this section we shall prove two lemmas which will enable us to reduce the general problem of large sample inference to the case where the variates under consideration have a joint normal distribution.

LEMMA 1. *For each positive integer n there exists a set-function $W_n^*(W_n)$ defined over all Borel measurable subsets W_n of the rn-dimensional sample space such that the following two conditions are fulfilled:*

(a) *For each W_n, $W_n^*(W_n)$ is a Borel measurable subset of the rn-dimensional sample space with the following property: For each point of $W_n^*(W_n)$ the maximum likelihood estimate exists and if a sample point E_n lies in $W_n^*(W_n)$ then also all those points E_n' lie in $W_n^*(W_n)$ for which $\hat{\theta}_n(E_n') = \hat{\theta}_n(E_n)$.*

(b) $\operatorname{Lim}_{n=\infty} \{P[W_n^*(W_n)|\theta] - P[W_n|\theta]\} = 0$ *uniformly in θ and W_n.*

Proof. Let λ be a real variable which takes only non-negative values and consider the region $W_n(\theta, \lambda)$ defined as the common part of the region W_n and the region $|n^{1/2}(\hat{\theta}_n - \theta)| \leqq \lambda$. Similarly let $W_n^*(W_n, \theta, \lambda)$ be the intersection of $W_n^*(W_n)$ and the region $|n^{1/2}(\hat{\theta}_n - \theta)| \leqq \lambda$.

For any function $\Phi(v)$ we will denote by g.l.b.$_v$ $\Phi(v)$ and l.u.b.$_v$ $\Phi(v)$ the greatest lower bound and the least upper bound of $\Phi(v)$ respectively.

Since, on account of Proposition I, for any sequence $\{\lambda_n\}$ for which $\lim_{n=\infty} \lambda_n = \infty$

$$\lim_{n=\infty} \{\text{g.l.b.}_\theta [P[|n^{1/2}(\hat{\theta}_n - \theta)| \leqq \lambda_n | \theta]]\} = 1,$$

Lemma I is proved if we show that there exists a sequence $\{\lambda_n\}$ not depending on θ and W_n such that $\lim_{n=\infty} \lambda_n = \infty$ and

$$\lim_{n=\infty} \{P[W_n(\theta, \lambda_n) | \theta] - P[W_n^*(W_n, \theta, \lambda_n) | \theta]\} = 0 \tag{20}$$

uniformly in W_n and θ.

Let q be a real variable restricted to values greater than 1. For any set of k integers $(r_1, \cdots, r_k)$ and for any value q denote by $I_n(r_1, \cdots, r_k, q)$ the region defined by the inequalities:

$$\frac{r_1 - 1/2}{qn^{1/2}} < \theta_n^1 < \frac{r_1 + 1/2}{qn^{1/2}}, \cdots, \frac{r_k - 1/2}{qn^{1/2}} < \theta_n^k < \frac{r_k + 1/2}{qn^{1/2}}. \tag{21}$$

Furthermore denote by $\theta_n(r_1, \cdots, r_k, q)$ the parameter point with the coordinates $r_1/qn^{1/2}, \cdots, r_k/qn^{1/2}$. We order the system of all sets of k integers $(r_1, \cdots, r_k)$ in a sequence and we shall denote by $I_{ns}(q)$ the interval $I_n(r_1, \cdots, r_k, q)$ where $(r_1, \cdots, r_k)$ is the sth element in the ordered sequence $(s = 1, 2, \cdots, \text{ad inf.})$. Similarly $\theta_{ns}(q)$ denotes the parameter point $\theta_n(r_1, \cdots, r_k, q)$ where $(r_1, \cdots, r_k)$ is the sth element in the ordered sequence.

Let $J_{ns}(W_n, q)$ be the common part of the three regions $I_{ns}(q)$, W_n and $Q_n[\theta_{ns}(q)]$ where for any θ, $Q_n(\theta)$ is defined by the inequalities

$$(22)\qquad \left|\frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta, \frac{1}{n^{1/3}}\right) + c_{ij}(\theta)\right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n),$$
$$\left|\frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta, \frac{1}{n^{1/3}}\right) + c_{ij}(\theta)\right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n).$$

The expression $\nu(n)$ is equal to l.u.b.$_\theta$ $[E_\theta\phi_{ij}(x, \theta, 2/n^{1/3}) - E_\theta\psi_{ij}(x, \theta, 2/n^{1/3})]$ and the expression $\bar{\nu}(n)$ is defined as follows: Denote by $c_{ij}(\theta, n)$ the least upper bound of $|c_{ij}(\theta) - c_{ij}(\bar{\theta})|$ with respect to $\bar{\theta}$ where $\bar{\theta}$ can take only values in the interval $[\theta - 1/n^{1/3}, \theta + 1/n^{1/3}]$. Then $\bar{\nu}(n)$ is defined as the least upper bound of $c_{ij}(\theta, n)$ with respect to θ, i and j. Because of condition (a) of Assumption III, we obtain

$$(23)\qquad \lim_{n=\infty}\nu(n) = \lim_{n=\infty}\bar{\nu}(n) = 0.$$

Let $J^*_{ns}(W_n, q)$ be a subset of $I_{ns}(q)$ such that the following two conditions are fulfilled:

(24a) If E_n is an element of $J^*_{ns}(W_n, q)$ then also all those points E_n' for which $\hat{\theta}_n(E_n') = \hat{\theta}_n(E_n)$ are elements of $J^*_{ns}(W_n, q)$, that is $J^*_{ns}(W_n, q)$ can be represented as a subset of the space of the maximum likelihood estimates. Furthermore $J^*_{ns}(W_n, q)$ is an interval in this space.

(24b) $\operatorname{Lim}_{n=\infty}\{\text{l.u.b.}_{s, W_n}|P[J_{ns}(W_n, q)|\theta_{ns}(q)] - \mathfrak{P}[J^*_{ns}(W_n, q)|\theta_{ns}(q)]|\} = 0$ where $\mathfrak{P}(V_n|\theta)$ denotes the probability of V_n calculated on the basis that the joint distribution of $n^{1/2}(\hat{\theta}_n^1 - \theta^1), \cdots, n^{1/2}(\hat{\theta}_n^k - \theta^k)$ is normal with zero means and covariance matrix $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$.

The existence of such a set $J^*_{ns}(W_n, q)$ can be proved as follows: Obviously there exists a subset $J^*_{ns}(W_r, q)$ of $I_{ns}(q)$ such that (24a) is fulfilled and

$$\mathfrak{P}[J^*_{ns}(W_n, q)\,|\,\theta_{ns}(q)] = \min\{P[J_{ns}(W_n, q)\,|\,\theta_{ns}(q)], \mathfrak{P}[I_{ns}(q)\,|\,\theta_{ns}(q)]\}.$$

Since $J_{ns}(W_n, q)$ is a subset of $I_{ns}(q)$ and since $\lim_{n=\infty}\{P[I_{ns}(q)|\theta_{ns}(q)] - \mathfrak{P}[I_{ns}(q)|\theta_{ns}(q)]\} = 0$ uniformly in s, the above defined subset $J^*_{ns}(W_n, q)$ satisfies also the condition (24b). We define

$$(25)\qquad W^*_n(W_n, q) = \sum_{s=1}^{\infty} J^*_{ns}(W_n, q).$$

Furthermore we define the regions $J_n(W_n, \theta, \lambda, q)$ and $J^*_n(W_n, \theta, \lambda, q)$ as follows:

$$(26)\qquad J_n(W_n, \theta, \lambda, q) = \sum_s J_{ns}(W_n, q),$$

$$(27)\qquad J^*_n(W_n, \theta, \lambda, q) = \sum_s J^*_{ns}(W_n, q).$$

where the summation is to be taken over all values of s for which $|\theta-\theta_{ns}(q)| \leqq \lambda/n^{1/2}$.

Let $\overline{J}^*_{ns}(W_n, q)$ be the intersection of $J^*_{ns}(W_n, q)$ and $Q_n[\theta_{ns}(q)]$, and let $\overline{J}^*_n(W_n, \theta, \lambda, q) = \sum_s \overline{J}^*_{ns}(W_n, q)$ where the summation is to be taken over all values of s for which $|\theta-\theta_{ns}(q)| \leqq \lambda/n^{1/2}$. Furthermore let $W^*_n(W_n, \theta, \lambda, q)$ be the intersection of $W^*_n(W_n, q)$ with the region $|n^{1/2}(\theta_n-\theta)| \leqq \lambda$.

If for a value of s we have $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ then for all points of $I_{ns}(q)$ we have $n^{1/2}|\theta-\theta_n| \leqq \lambda+1/q$. If there exists a point in $I_{ns}(q)$ for which $n^{1/2}|\theta-\theta_n| \leqq \lambda-1/q$ then $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$. Hence $W^*_n(W_n, \theta, \lambda-1/q, q)$ is a subset of $J^*_n(W_n, \theta, \lambda, q)$, and the latter is a subset of $W^*_n(W_n, \theta, \lambda+1/q, q)$. Thus

$$
\begin{aligned}
&|P[W^*_n(W_n, \theta, \lambda, q)\,|\,\theta] - P[J^*_n(W_n, \theta, \lambda, q)\,|\,\theta]| \\
&\leqq P[W^*_n(W_n, \theta, \lambda+1/q, q)\,|\,\theta] - P[W^*_n(W_n, \theta, \lambda-1/q, q)\,|\,\theta] \\
&\leqq P[\lambda - 1/q \leqq n^{1/2}|\theta_n - \theta| \leqq \lambda + 1/q\,|\,\theta].
\end{aligned} \tag{28}
$$

According to condition (b) of Assumption III $E_\theta[\phi_{ij}(x, \theta, \delta)]^2$ and $E_\theta[\psi_{ij}(x, \theta, \delta)]^2$ are bounded functions of θ. Hence also $E_\theta\phi_{ij}(x, \theta, \delta)$ and $E_\theta\psi_{ij}(x, \theta, \delta)$ are bounded functions of θ. Substituting $\delta=0$ we find that $c_{ij}(\theta)$ is a bounded function of θ. From the boundedness of $c_{ij}(\theta)$ and from the fact that the determinant $|c_{ij}(\theta)|$ has a positive lower bound (condition (c) of Assumption III) it follows because of Proposition I that

$$
\lim_{q=\infty} \{\text{l.u.b.}_\theta\, P[\lambda - 1/q \leqq n^{1/2}|\theta_n - \theta| \leqq \lambda + 1/q\,|\,\theta]\} = 0 \tag{29}
$$

uniformly in λ. From (28) and (29) we obtain

$$
\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W^*_n(W_n, \theta, \lambda, q)\,|\,\theta] - P[J^*_n(W_n, \theta, \lambda, q)\,|\,\theta]|\} = \epsilon_1(\lambda, q) \tag{30}
$$

where $\lim_{q=\infty} \epsilon_1(\lambda, q) = 0$ uniformly in λ.

Denote by $R_n(\theta, \lambda, q)$ the common part of the regions $Q_n[\theta_{ns}(q)]$ formed for all values s for which $|\theta-\theta_{ns}(q)| \leqq \lambda/n^{1/2}$. Then for almost all n the region $R_n(\theta, \lambda, q)$ contains the region $T_n(\theta)$ as a subset, where $T_n(\theta)$ is defined by the inequalities

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) + c_{ij}(\theta)\right| < \frac{1}{n^{1/3}} + \nu(n), \\
&\left|\frac{1}{n}\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) + c_{ij}(\theta)\right| < \frac{1}{n^{1/3}} + \nu(n).
\end{aligned} \tag{31}
$$

In fact, from (31) it follows that

$$
\begin{aligned}
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n), \\
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n).
\end{aligned} \tag{32}
$$

Since

$$
\begin{aligned}
\frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) &\leqq \frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq \frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) &\leqq \frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq \frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right)
\end{aligned}
$$

for almost all n and for all s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$, we obtain from (32)

$$
\begin{aligned}
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n), \\
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n),
\end{aligned} \tag{33}
$$

for almost all n and for all s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$. Since $|c_{ij}[\theta_{ns}(q)] - c_{ij}(\theta)| \leqq \nu(n)$ for almost all n and for all s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$, we obtain from (33)

$$
\begin{aligned}
c_{ij}[\theta_{ns}(q)] - \frac{1}{n^{1/3}} - \nu(n) - \bar{\nu}(n) &\leqq \frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq c_{ij}[\theta_{ns}(q)] + \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n), \\
c_{ij}[\theta_{ns}(q)] - \frac{1}{n^{1/3}} - \nu(n) - \bar{\nu}(n) &\leqq \frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq c_{ij}[\theta_{ns}(q)] + \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n)
\end{aligned} \tag{34}
$$

for almost all n and for all s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$. The inequalities (34) are equivalent to the inequalities (22) if in (22) $\theta_{ns}(q)$ is substituted for θ. Hence our statement about the region $R_n(\theta, \lambda, q)$ is proved.

Consider the region $U_n(\theta)$ defined by the inequalities

$$\left|\frac{1}{n}\sum_{\alpha}\phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) - E_\theta\phi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right)\right| < \frac{1}{n^{1/3}},$$
$$\left|\frac{1}{n}\sum_{\alpha}\psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) - E_\theta\psi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right)\right| < \frac{1}{n^{1/3}}. \tag{35}$$

Since

$$\nu(n) \geqq \left|E_\theta\phi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right) + c_{ij}(\theta)\right|$$

and

$$\nu(n) \geqq \left|c_{ij}(\theta) + E_\theta\psi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right)\right|,$$

the validity of (35) implies the validity of (31). Hence $U_n(\theta)$ is a subset of $T_n(\theta)$. From condition (b) of Assumption III and Tshebysheff's inequality it follows that $\lim_{n=\infty} P[U_n(\theta)\,|\,\theta] = 1$ uniformly in θ. Hence $\lim_{n=\infty} P[T_n(\theta)\,|\,\theta] = 1$ uniformly in θ. Thus, as can easily be seen,

$$\lim_{n=\infty} P[R_n(\theta, \lambda_n, q)\,|\,\theta] = 1 \tag{36}$$

for any bounded sequence $\{\lambda_n\}$ uniformly in θ and q.

Let $\dot{J}_{ns}(W_n, q)$ be the intersection of the regions W_n and $I_{ns}(q)$. Furthermore let $\dot{J}_n(W_n, \theta, \lambda, q)$ be equal to $\sum_s \dot{J}_{ns}(W_n, q)$ where the summation is to be taken over all values of s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$. Then the common part of $\sum_{s=1}^{\infty} I_{ns}(q)$ and $W_n(\theta, \lambda - 1/q)$ is a subset of $\dot{J}_n(W_n, \theta, \lambda, q)$, and the last is a subset of the common part of $\sum_{s=1}^{\infty} I_{ns}(q)$ and $W_n(\theta, \lambda + 1/q)$. Hence, since $P[\sum_{s=1}^{\infty} I_{ns}(q)\,|\,\theta] = 1$, we have

$$\begin{aligned} &|P[W_n(\theta, \lambda)\,|\,\theta] - P[\dot{J}_n(W_n, \theta, \lambda, q)\,|\,\theta]| \\ &\qquad \leqq P[W_n(\theta, \lambda + 1/q)\,|\,\theta] - P[W_n(\theta, \lambda - 1/q)\,|\,\theta] \\ &\qquad \leqq P[\lambda - 1/q \leqq n^{1/2}|\theta_n - \theta| \leqq \lambda + 1/q\,|\,\theta]. \end{aligned} \tag{37}$$

From (37) and (29) it follows that for any sequence $\{q_n\}$ for which $\lim q_n = \infty$, we have

$$\lim_{n=\infty} |P[W_n(\theta, \lambda)\,|\,\theta] - P[\dot{J}_n(W_n, \theta, \lambda, q_n)\,|\,\theta]| = 0 \tag{38}$$

uniformly in θ, W_n and λ.

Since the common part of the regions $\dot{J}_n(W_n, \theta, \lambda, q)$ and $R_n(\theta, \lambda, q)$ is contained as a subset in $J_n(W_n, \theta, \lambda, q)$ and since the last is a subset of $\dot{J}_n(W_n, \theta, \lambda, q)$, we obtain from (38) and (36) that for any bounded sequence

$\{\lambda_n\}$ and for any sequence $\{q_n\}$ for which $\lim_{n=\infty} q_n = \infty$ we have

$$\lim_{n=\infty} \{P[W_n(\theta, \lambda_n) \mid \theta] - P[J_n(W_n, \theta, \lambda_n, q_n) \mid \theta]\} = 0 \tag{39}$$

uniformly in θ and W_n.

Since the common part of the regions $J_n^*(W_n, \theta, \lambda, q)$ and $R_n(\theta, \lambda, q)$ is a subset of $\overline{J}_n^*(W_n, \theta, \lambda, q)$ and the last is a subset of $J_n^*(W_n, \theta, \lambda, q)$ we obtain from (36) that for any bounded sequence $\{\lambda_n\}$ and for any sequence $\{q_n\}$ for which $\lim_{n=\infty} q_n = \infty$, we have

$$\lim_{n=\infty} \{P[\overline{J}_n^*(W_n, \theta, \lambda_n, q_n) \mid \theta] - P[J_n^*(W_n, \theta, \lambda_n, q_n) \mid \theta]\} = 0 \tag{40}$$

uniformly in θ and W_n.

Now we shall evaluate the limit values of $P[J_n(W_n, \theta, \lambda, q) \mid \theta]$ and $P[\overline{J}_n^*(W_n, \theta, \lambda, q) \mid \theta]$. Denote by $A_n(\lambda, q)$ the domain in the space of the variables θ, W_n, θ' and E_n defined as follows: θ and W_n can take arbitrary values, θ' is restricted to values for which $|\theta' - \theta| \leqq \lambda/n^{1/2}$; and for any θ and W_n, E_n is restricted to points which lie in the sum of the sets $J_n(W_n, \theta, \lambda, q)$ and $\overline{J}_n^*(W_n, \theta, \lambda, q)$. Denote furthermore by $\rho_n(\theta', \theta, \hat{\theta}_n)$ the function

$$\rho_n(\theta', \theta, \hat{\theta}_n) = -\frac{1}{2} \sum_j \sum_i n(\theta'^i - \hat{\theta}_n^i)(\theta'^j - \hat{\theta}_n^j) c_{ij}(\theta). \tag{41}$$

Consider the Taylor expansion

$$\begin{aligned} &\sum_\alpha \log f(x_\alpha, \theta') \\ &= \sum_\alpha \log f(x_\alpha, \hat{\theta}_n) + \frac{n}{2} \sum_j \sum_i (\theta'^i - \hat{\theta}_n^i)(\theta'^j - \hat{\theta}_n^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} \end{aligned} \tag{42}$$

where $\bar{\theta}$ lies in the interval $[\theta', \hat{\theta}_n]$.

Since in the domain $A_n(\lambda, q)$ any point E_n lies in the sum of the sets $J_n(W_n, \theta, \lambda, q)$ and $\overline{J}_n^*(W_n, \theta, \lambda, q)$, it follows from the definitions of these sets that E_n lies in the set $\sum_s Q_n[\theta_{ns}(q)]$ where the summation is to be taken over all values of s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$ (the set $Q_n(\theta)$ is defined in (22)). Hence for any E_n in the domain $A_n(\lambda, q)$ we have

$$\begin{aligned} &\left| \frac{1}{n} \sum_\alpha \phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) + c_{ij}[\theta_{ns}(q)] \right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \nu(n), \\ &\left| \frac{1}{n} \sum_\alpha \psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) + c_{ij}[\theta_{ns}(q)] \right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \nu(n) \end{aligned} \tag{43}$$

for that value of s for which E_n lies in $I_{ns}(q)$. In all that follows, with any

point of the domain $A_n(\lambda, q)$ we shall associate the integer s for which E_n lies in $I_{ns}(q)$. Since $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ in the domain $A_n(\lambda, q)$, we have in the domain $A_n(\lambda, q)$

$$\frac{1}{n}\sum_\alpha \psi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right] \leqq \frac{1}{n}\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \leqq \frac{1}{n}\sum_\alpha \phi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right],$$

(44)

$$\frac{1}{n}\sum_\alpha \psi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right] \leqq \frac{1}{n}\sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \leqq \frac{1}{n}\sum_\alpha \phi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right]$$

for almost all values of n. From the definition of $\bar{\nu}(n)$ and from the validity of the inequality $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ in $A_n(\lambda, q)$ we find that in the domain $A_n(\lambda, q)$

$$|c_{ij}[\theta_{ns}(q)] - c_{ij}(\theta)| \leqq \bar{\nu}(n) \tag{45}$$

for almost all values of n. From (43), (44) and (45) it follows that in the domain $A_n(\lambda, q)$

$$\left|\frac{1}{n}\sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) + c_{ij}(\theta)\right| \leqq \frac{1}{n^{1/3}} + \nu(n) + 2\bar{\nu}(n),$$

(46)

$$\left|\frac{1}{n}\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) + c_{ij}(\theta)\right| \leqq \frac{1}{n^{1/3}} + \nu(n) + 2\bar{\nu}(n)$$

for almost all values of n.

Since $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ in $A_n(\lambda, q)$, we have

$$n^{1/2}|\theta - \theta_n| \leqq \lambda + 1/q \quad \text{in} \quad A_n(\lambda, q) \tag{47}$$

and therefore

$$n^{1/2}|\theta' - \theta_n| \leqq 2\lambda + 1/q \quad \text{in} \quad A_n(\lambda, q). \tag{48}$$

Since $\bar{\theta}$ lies in the interval $[\theta', \theta_n]$, from (48) we obtain

$$n^{1/2}|\bar{\theta} - \theta_n| \leqq 2\lambda + 1/q \quad \text{in} \quad A_n(\lambda, q). \tag{49}$$

From (47) and (49) it follows that in $A_n(\lambda, q)$

$$\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \leqq \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} \leqq \sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \tag{50}$$

for almost all n. Because of (23), (46) and (50) we obtain from (41) and (42)

$$\lim_{n=\infty}\left\{\text{l.u.b.}_{A_n(\lambda,q)}\left|\sum_a \log f(x_a,\theta')-\sum_a \log f(x_a,\hat{\theta}_n)-\rho_n(\theta',\theta,\hat{\theta}_n)\right|\right\}=0, \tag{51}$$

and

$$\lim_{n=\infty}\left\{\text{l.u.b.}_{A_n(\lambda,q)}\left|\left[\sum_a \log f(x_a,\theta)-\sum_a \log f(x_a,\theta')\right]-[\rho_n(\theta,\theta,\hat{\theta}_n)-\rho_n(\theta',\theta,\hat{\theta}_n)]\right|\right\}=0. \tag{52}$$

Denote $P[J_{ns}(W_n,q)\,|\,\theta_{ns}(q)]$ by $P_{ns}(W_n,q)$ and $P[\bar{J}^*_{ns}(W_n,q)\,|\,\theta_{ns}(q)]$ by $\bar{P}^*_{ns}(W_n,q)$. Substituting $\theta_{ns}(q)$ for θ' we easily obtain from (52)

$$\lim_{n=\infty}\left\{\text{l.u.b.}_{\theta,W_n}\left|P[J_n(W_n,\theta,\lambda,q)\,|\,\theta]-\sum_s P_{ns}(W_n,q)\exp\left(\rho_n[\theta,\theta,\theta^*_{ns}(q)]-\rho_n[\theta_{ns}(q),\theta,\theta^*_{ns}(q)]\right)\right|\right\}=0 \tag{53}$$

where the summation with respect to s is to be taken over all values for which $n^{1/2}|\theta-\theta_{ns}(q)|\leqq\lambda$ and $\theta^*_{ns}(q)$ is a parameter point for which

$$n^{1/2}\left|\theta_{ns}(q)-\theta^*_{ns}(q)\right|\leqq 1/q. \tag{54}$$

Since $c_{ij}(\theta)$ is a uniformly continuous and bounded function of θ, it follows from (41) and (54) that

$$\begin{aligned}\left|\rho(\theta,\theta,\theta^*_{ns}(q))-\rho[\theta,\theta,\theta_{ns}(q)]\right|&=\phi_1[\theta,\theta^*_{ns}(q),\theta_{ns}(q)]/q,\\ \left|\rho[\theta_{ns}(q),\theta,\theta^*_{ns}(q)]-\rho[\theta_{ns}(q),\theta,\theta_{ns}(q)]\right|&=\phi_2[\theta,\theta^*_{ns}(q),\theta_{ns}(q)]/q,\end{aligned} \tag{55}$$

where $\phi_1[\theta,\theta^*_{ns}(q),\theta_{ns}(q)]$ and $\phi_2[\theta,\theta^*_{ns}(q),\theta_{ns}(q)]$ are bounded functions of θ, $\theta^*_{ns}(q)$ and $\theta_{ns}(q)$ in the domain $n^{1/2}|\theta-\theta_{ns}(q)|\leqq\lambda$. From (53) and (55) it follows that

$$\limsup_{n=\infty}\left\{\text{l.u.b.}_{\theta,W_n}\left|P[J_n(W_n,\theta,\lambda,q)\,|\,\theta]-\sum_s P_{ns}(W_n,q)\exp\left(\rho_n[\theta,\theta,\theta_{ns}(q)]-\rho_n[\theta_{ns}(q),\theta,\theta_{ns}(q)]\right)\right|\right\}=\epsilon(\lambda,q) \tag{56}$$

where

$$\lim_{q=\infty}\epsilon(\lambda,q)=0 \tag{57}$$

uniformly in λ over any finite positive interval. Similarly we obtain

$$\limsup_{n\to\infty}\left\{\text{l.u.b.}_{\theta,W_n}\left|P[J_n^*(W_n,\theta,\lambda,q)\,|\,\theta]-\sum_s P_{ns}^*(W_n,q)\exp\left(\rho_n[\theta,\theta,\theta_{ns}(q)]-\rho_n[\theta_{ns}(q),\theta,\theta_{ns}(q)]\right)\right|\right\}=\eta(\lambda,q) \tag{58}$$

where

$$\lim_{q\to\infty}\eta(\lambda,q)=0 \tag{59}$$

uniformly in λ over any finite positive interval. In the formulas (56) and (58) the summation with respect to s is to be taken for all values for which $|\theta-\theta_{ns}(q)|\leqq\lambda/n^{1/2}$. The expression $\rho_n[\theta_{ns}(q),\theta,\theta_{ns}(q)]$ is obviously equal to zero, hence (56) and (58) can be simplified by substituting zero for this expression. Denote $P[J_{ns}^*(W_n,q)\,|\,\theta_{ns}(q)]$ by $P_{ns}^*(W_n,q)$. Because of Proposition I we have

$$\lim_{n\to\infty}\{\mathfrak{P}[J_{ns}^*(W_n,q)\,|\,\theta_{ns}(q)]-P_{ns}^*(W_n,q)\}=0$$

uniformly in s and W_n. Hence we obtain from condition (b) of (24)

$$\lim_{n\to\infty}\{P_{ns}(W_n,q)-P_{ns}^*(W_n,q)\}=0 \tag{60}$$

uniformly in s and W_n. Since $\mathfrak{J}_{ns}^*(W_n,q)$ is the intersection of $Q_n[\theta_{ns}(q)]$ with $J_{ns}^*(W_n,q)$ and since

$$\lim_{n\to\infty}P[Q_n[\theta_{ns}(q)]\,|\,\theta_{ns}(q)]=1$$

uniformly in s, we have

$$\lim_{n\to\infty}\{\bar{P}_{ns}^*(W_n,q)-P_{ns}^*(W_n,q)\}=0 \tag{61}$$

uniformly in s and W_n. Since for any given λ and q the number of different values of s satisfying the inequality $n^{1/2}|\theta-\theta_{ns}(q)|\leqq\lambda$ is a bounded function of θ, from (56), (58), (60) and (61) we obtain

$$\limsup_{n\to\infty}\{\text{l.u.b.}_{\theta,W_n}|P[J_n(W_n,\theta,\lambda,q)\,|\,\theta]-P[J_n^*(W_n,\theta,\lambda,q)\,|\,\theta]|\}=\zeta(\lambda,q)\leqq\epsilon(\lambda,q)+\eta(\lambda,q). \tag{62}$$

Hence

$$\lim_{q\to\infty}\zeta(\lambda,q)=0 \tag{63}$$

uniformly in λ over any finite positive interval.

For any positive λ' the sets $J_n(W_n, \theta, \lambda+\lambda', q)-J_n(W_n, \theta, \lambda, q)$, $\overline{J}_n^*(W_n, \theta, \lambda+\lambda', q)-\overline{J}_n^*(W_n, \theta, \lambda, q)$, $J_n^*(W_n, \theta, \lambda+\lambda', q)-J^*(W_n, \theta, \lambda, q)$, $W_n^*(W_n, \theta, \lambda+\lambda', q)-W_n^*(W_n, \theta, \lambda, q)$ and $W_n(\theta, \lambda+\lambda')-W_n(\theta, \lambda)$ are subsets of the set defined by the inequality

$$\lambda - 1/q \leqq n^{1/2}|\theta - \theta_n| \leqq \lambda + \lambda' + 1/q.$$

Since for any sequence $\{\lambda_n\}$ for which $\lim_{n=\infty} \lambda_n = \infty$ we have

$$\lim_{n=\infty} P[\lambda_n - 1/q \leqq n^{1/2}|\theta - \theta_n| \leqq \lambda_n + \lambda' + 1/q \,|\, \theta] = 0$$

uniformly in θ, q and λ', (39) and (40) hold for any arbitrary sequence $\{\lambda_n\}$ and (63) holds uniformly in λ where λ can take any positive value. Thus from (30), (39), (40), (62) and (63) we obtain

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda)\,|\,\theta] - P[W_n^*(W_n, \theta, \lambda, q)\,|\,\theta]|\} = \epsilon_3(\lambda, q), \tag{64}$$

where

$$\lim_{q=\infty} \epsilon_3(\lambda, q) = 0 \tag{65}$$

uniformly in λ. Let $\{q_i\}$ $(i=1, 2, \cdots, \text{ad inf.})$ be a sequence of positive integers such that $\lim_{i=\infty} q_i = +\infty$. Furthermore let $\{\eta_i\}$ be a sequence of positive numbers such that $\lim_{i=\infty} \eta_i = 0$. We define $W_n^*(W_n)$ as follows:

$$W_n^*(W_n) = W_n^*(W_n, q_{i+1}) \quad \text{for} \quad n_i < n \leqq n_{i+1} \quad (i = 0, 1, \cdots, \text{ad inf.}). \tag{66}$$

The sequence $\{n_i\}$ $(i=0, 1, 2, \cdots, \text{ad inf.})$ of integers is chosen as follows: Denote by $F_n(\lambda, q)$ the expression

$$\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda)\,|\,\theta] - P[W_n^*(W_n, \theta, \lambda, q)\,|\,\theta]|.$$

The integer n_0 is put equal to 0 and n_i is chosen such that

$$\begin{gathered} n_i > n_{i-1}, \\ F_n(\lambda_i, q_{i+1}) < \epsilon_3(\lambda_i, q_{i+1}) + \eta_i \end{gathered} \tag{67}$$

for all $n > n_i$, and where $\{\lambda_i\}$ denotes a sequence of numbers such that $\lim_{i=\infty} \lambda_i = +\infty$.

Let $\lambda_n' = \lambda_i$, $\eta_n' = \eta_i$, and $q_n' = q_{i+1}$ for $n_i < n \leqq n_{i+1}$ $(i=0, 1, 2, \cdots, \text{ad inf.})$. Then from (64), (65) and (67) we obtain

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda_n')\,|\,\theta] - P[W_n^*(W_n, \theta, \lambda_n', q_n')\,|\,\theta]|\} = 0. \tag{68}$$

Denote by $W_n^*(W_n, \theta, \lambda)$ the intersection of $W_n^*(W_n)$ and the set defined by the inequality $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$. Since $W_n^*(W_n, \theta, \lambda, q)$ is the intersection of $W_n^*(W_n, q)$ and the region $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$, it follows from (66) that

$$W_n^*(W_n, \theta, \lambda_n', q_n') = W_n^*(W_n, \theta, \lambda_n'). \tag{69}$$

Equation (20) follows from (68) and (69). Hence Lemma 1 is proved.

We shall say that a region V_n^* lies in the space of the maximum likelihood estimates if it has the following property: If E_n is an element of V_n^* then also all those points E_n' for which $\hat{\theta}_n(E_n') = \hat{\theta}_n(E_n)$ are elements of V_n^*. In all the following considerations the symbol * as a superscript in the notation of a region will indicate that the region lies in the space of the maximum likelihood estimates, except if a statement to the contrary is explicitly made. For any region V_n^* we shall denote by $\mathfrak{P}(V_n^*|\theta)$ the probability that the sample point will fall in V_n^* calculated under the assumption that $n^{1/2}(\hat{\theta}_n^1 - \theta^1), \cdots, n^{1/2}(\hat{\theta}_n^k - \theta^k)$ have a joint normal distribution with zero means and covariance matrix $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$.

LEMMA 2. *There exists a function $W_n^*(R_n^*)$ defined over all Borel measurable subsets R_n^* such that*

$$\lim_{n=\infty} \{P[R_n^*|\theta] - \mathfrak{P}[W_n^*(R_n^*)|\theta]\} = 0$$

uniformly in θ and R_n^*.

Proof. Since we assumed that the set $J_{ns}^*(W_n, q)$ defined in (24) is an interval in the space of the maximum likelihood estimates, it follows from Proposition I that

$$\lim_{n=\infty} \{P[J_{ns}^*(W_n, q)|\theta] - \mathfrak{P}[J_{ns}^*(W_n, q)|\theta]\} = 0 \tag{70}$$

uniformly in θ, W_n, and s. Let $W_n^*(W_n, q)$ be the set defined in (25) and let $W_n^*(W_n, \theta, \lambda, q)$ be the intersection of $W_n^*(W_n, q)$ and the region $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$. For given values of λ and q the number of different values of s, for which $J_{ns}^*(W_n, q)$ has at least a point common with the region $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$, is a bounded function of θ. Hence it follows from (70) that

$$\lim_{n=\infty} \{P[W_n^*(W_n, \theta, \lambda, q)|\theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda, q)|\theta]\} = 0 \tag{71}$$

uniformly in θ and W_n. From (64), (65) and (71) we obtain

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda)|\theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda, q)|\theta]|\} = \epsilon(\lambda, q) \tag{72}$$

where

$$\lim_{q=\infty} \epsilon(\lambda, q) = 0 \tag{73}$$

uniformly in λ. The set $W_n(\theta, \lambda)$ denotes the intersection of W_n and the region $n^{1/2}|\theta - \theta_n| \leqq \lambda$.

Let $\{q_i\}$ $(i=1, 2, \cdots, \text{ad inf.})$ be a sequence of positive integers such that $\lim_{i=\infty} q_i = \infty$. Furthermore let $\{\eta_i\}$ be a sequence of positive numbers such that $\lim_{i=\infty} \eta_i = 0$. We define $W_n^*(W_n)$ as follows

$$W_n^*(W_n) = W_n^*(W_n, q_{i+1}) \text{ for } n_i < n \leqq n_{i+1} \quad (i = 0, 1, 2, \cdots, \text{ad inf.}). \tag{74}$$

The sequence $\{n_i\}$ $(i=0, 1, 2, \cdots, \text{ad inf.})$ of integers is chosen as follows: Denote by $F_n(\lambda, q)$ the expression

$$\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda) | \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda, q) | \theta]|.$$

The integer n_0 is put equal to zero and n_i is chosen so that

$$\begin{gathered} n_i > n_{i-1}, \\ F_n(\lambda_i, q_{i+1}) < \epsilon(\lambda_i, q_{i+1}) + \eta_i \end{gathered} \tag{75}$$

for all $n > n_i$, and $\{\lambda_i\}$ denotes a sequence of numbers such that $\lim \lambda_i = \infty$. Let $\lambda_n' = \lambda_i$, $\eta_n' = \eta_i$ and $q_n' = q_{i+1}$ for $n_i < n \leqq n_{i+1}$ $(i=0, 1, \cdots, \text{ad inf.})$. From (72), (73) and (75) we obtain

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda_n') | \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda_n', q_n') | \theta]|\} = 0. \tag{76}$$

Denote by $W_n^*(W_n, \theta, \lambda)$ the intersection of $W_n^*(W_n)$ and the region $n^{1/2}|\theta - \theta_n| \leqq \lambda$. Because of (74) we obviously have

$$W_n^*(W_n, \theta, \lambda_n') = W_n^*(W_n, \theta, \lambda_n', q_n').$$

Hence from (76) we obtain

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda_n') | \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda_n') | \theta]|\} = 0. \tag{77}$$

Since $\lim_{n=\infty} \lambda_n' = \infty$, it follows from (77) that

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n | \theta] - \mathfrak{P}[W_n^*(W_n) | \theta]|\} = 0. \tag{78}$$

The region W_n may be any Borel measurable subset of the rn-dimensional sample space. In particular, W_n may be any Borel measurable subset R_n^* in the space of the maximum likelihood estimates. Hence Lemma 2 follows from (78).

On the basis of Lemmas 1 and 2 we can restrict ourselves in case of large

samples to subsets of the space of the maximum likelihood estimates and we can substitute $\mathfrak{P}[V_n^*|\theta]$ for $P[V_n^*|\theta]$. Hence, if the sample is sufficiently large, the problem of statistical inference concerning the unknown parameter θ can be reduced to the case where the variates involved have a joint normal distribution.

5. **Tests of simple hypotheses which have uniformly best average power over a family of surfaces.** For any value c let K_c denote a surface in the parameter space. For instance K_c may be defined by the equation $\phi(\theta)=c$ where $\phi(\theta)$ denotes some analytic function of θ. Consider furthermore a non-negative function $w(\theta)$ of θ, called a weight function. For any function $\psi(\theta)$ of θ the symbol $\int_{K_c}\psi(\theta)dA$ will denote the surface integral of the function $\psi(\theta)$ over the surface K_c.

DEFINITION I. *A critical region W_n is said to have uniformly best average power with respect to the surfaces K_c and the weight function $w(\theta)$ if for any region Z_n of size equal to that of W_n we have $\int_{K_c}P(W_n|\theta)w(\theta)dA \geqq \int_{K_c}P(Z_n|\theta)w(\theta)dA$ for all values c for which K_c is defined.*

Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution. The mean values $\theta^1, \cdots, \theta^k$ of the variates $y^1, \cdots, y^k$ are unknown, but the covariance matrix $\|\sigma_{ij}\|$ $(i, j=1, \cdots, k)$ is known and is nonsingular. Suppose that we wish to test the simple hypothesis that $\theta=\theta_0$. Consider the family of ellipsoids given by

$$\sum_j \sum_i \lambda_{ij}[\theta^i - \theta_0^i][\theta^j - \theta_0^j] = c, \tag{79}$$

where $\|\lambda_{ij}\| = \|\sigma_{ij}\|^{-1}$. For any c denote by S_c the ellipsoid given by (79). Consider a nonsingular linear transformation of the parameter space

$$\theta'^i - \theta_0^i = \beta_{i1}(\theta^1 - \theta_0^1) + \cdots + \beta_{ik}(\theta^k - \theta_0^k) \tag{80}$$

such that the family of ellipsoids S_c is transformed into a family of concentric spheres with the center at θ_0. Denote by S_c' the image of S_c. For any point θ and for any positive ρ consider the set $\omega(\theta, \rho)$ consisting of all points θ_1 which lie on the same S_c as θ and for which $|\theta_1-\theta| \leqq \rho$. Let

$$\xi(\theta) = \lim_{\rho=0} \frac{A[\omega'(\theta, \rho)]}{A[\omega(\theta, \rho)]}, \tag{81}$$

where $\omega'(\theta, \rho)$ is the image of $\omega(\theta, \rho)$ and for any set ω, $A(\omega)$ denotes the area of ω.

PROPOSITION II. *If the variates $y^1, \cdots, y^k$ have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and a known covariance matrix $\|\sigma_{ij}\|$, then for testing the hypothesis $\theta=\theta_0$ on the basis of a single observation on each of the variates $y^1, \cdots, y^k$, the critical region given by the inequality*

$$\sum_j \sum_i \lambda_{ij}(y^i - \theta_0^i)(y^j - \theta_0^j) \geqq d \qquad (\|\lambda_{ij}\| = \|\sigma_{ij}\|^{-1}) \tag{82}$$

has uniformly best average power with respect to the surfaces S_c defined in (79) *and the weight function given in* (81).

Proof. Consider the linear transformation

$$y'^i - \theta_0^i = \beta_{i1}(y^1 - \theta_0^1) + \cdots + \beta_{ik}(y^k - \theta_0^k), \tag{83}$$

where the matrix $\|\beta_{ij}\|$ is the same as in (80). The variates y'^i $(i=1, \cdots, k)$ are normally and independently distributed with mean values $\theta_0^1, \cdots, \theta_0^k$ (under the hypothesis $\theta=\theta_0$) and have a common variance σ^2. We will assume $\sigma^2=1$, since this can always be achieved by multiplying the matrix $\|\beta_{ij}\|$ by a proportionality factor. The critical region W given in (82) will be transformed into the region W' given by

$$(y'^1 - \theta_0^1)^2 + \cdots + (y'^k - \theta_0^k)^2 \geqq d. \tag{84}$$

Because of (81) we obviously have

$$\int_{S_c} P(Z \mid \theta)\xi(\theta)dA = \int_{S_c'} P(Z' \mid \theta')dA, \tag{85}$$

where Z denotes an arbitrary region in the space of $y^1, \cdots, y^k$ and Z' is the image of Z in the space of $y'^1, \cdots, y'^k$. Hence in order to prove Proposition II we have merely to show that

$$\int_{S_c'} P(W' \mid \theta')dA \geqq \int_{S_c'} P(Z' \mid \theta')dA \tag{86}$$

for any region Z' in the space of $y'^1, \cdots, y'^k$ which has a size equal to that of W'.

By a lemma of Neyman and Pearson(4) we see easily that (86) is proved, if we can show that there exists a function $d(c)$ of c such that

$$\int_{S_c'} p(y' \mid \theta')dA / p(y' \mid \theta_0) \geqq d(c) \text{ within } W' \tag{87}$$

and

$$\int_{S_c'} p(y' \mid \theta')dA / p(y' \mid \theta_0) \leqq d(c) \text{ outside } W' \tag{88}$$

for all positive values of c, where $p(y' \mid \theta')$ denotes the joint density function of $y'^1, \cdots, y'^k$ under the assumption that the true means are $\theta'^1, \cdots, \theta'^k$.

(4) J. Neyman and E. S. Pearson, *Contributions to the theory of testing statistical hypotheses*, Statistical Research Memoirs vol. 1 (1936).

If we denote $y'^i - \theta_0^i$ by v^i and $\theta'^i - \theta_0^i$ by θ^{*i}, we have

$$p(y' \mid \theta') = \frac{1}{(2\pi)^{k/2}} \exp\left(-2^{-1}\sum (v^i - \theta^{*i})^2\right)$$

and

$$p(y' \mid \theta_0) = \frac{1}{(2\pi)^{k/2}} \exp\left(-2^{-1}\sum (v^i)^2\right).$$

Hence

$$\begin{aligned}\int_{S_c'} p(y' \mid \theta')dA &= \left(\frac{1}{2\pi}\right)^{k/2} \int_{S_c'} \exp\left(-2^{-1}\sum (v^i - \theta^{*i})^2\right)dA \\ &= \left(\frac{1}{2\pi}\right)^{k/2} \exp\left(-2^{-1}\sum (v^i)^2\right) \\ &\quad \cdot \int_{S_c'} \exp\left(\sum v^i\theta^{*i} - 2^{-1}\sum (\theta^{*i})^2\right)dA \\ &= p(y' \mid \theta_0) \exp\left(-2^{-1}\sum (\theta^{*i})^2\right) \int_{S_c'} \exp\left(\sum v^i\theta^{*i}\right)dA,\end{aligned}$$

since $\sum(\theta^{*i})^2$ is constant on the surface S_c'. Hence (87) and (88) can be written

$$I(v^1, \cdots, v^k) = \int_{S_c'} \exp \sum v^i\theta^{*i} dA \geqq d^*(c) \text{ within } W', \tag{89}$$

$$I(v^1, \cdots, v^k) \leqq d^*(c) \text{ outside } W'. \tag{90}$$

Denote $\left|(\sum(v^i)^2)^{1/2}\right|$ by r_v and $\left|(\sum(\theta^{*i})^2)^{1/2}\right|$ by r^*. On the surface S_c' we have $r^* = c$. Denote by $\alpha(\theta^*)$ the angle $(0 \leqq \alpha \leqq \pi)$ between the vector v and the vector θ^*. Then we have

$$I(v^1, \cdots, v^k) = \int_{S_c'} \exp\left(cr_v \cos\left[\alpha(\theta^*)\right]\right)dA.$$

Because of the symmetry of the sphere, the value of this integral will not be changed if we substitute $\beta(\theta^*)$ for $\alpha(\theta^*)$ where $\beta(\theta^*)$ denotes the angle $(0 \leqq \beta(\theta^*) \leqq \pi)$ between the vector θ^* and an arbitrarily chosen fixed vector u. Hence $I(v^1, \cdots, v^k)$ depends only on r_v, that is $I(v^1, \cdots, v^k) = I(r_v)$. The inequalities (89) and (90) are obviously proved if we can show that $I(r_v)$ is a monotonically increasing function of r_v. We have

$$\frac{dI(r_v)}{dr_v} = \int_{S_c'} c \cos\left[\beta(\theta^*)\right] \exp\left(cr_v \cos\left[\beta(\theta^*)\right]\right)dA. \tag{91}$$

Denote by ω_1 the subset of S_c' in which $0 \leqq \beta(\theta^*) \leqq \pi/2$, and by ω_2 the subset in which $\pi/2 \leqq \beta(\theta^*) \leqq \pi$. Because of the symmetry of the sphere we obvi-

ously have

$$\int_{\omega_2} c \cos [\beta(\theta^*)] \exp (cr_v \cos [\beta(\theta^*)])dA$$

$$= \int_{\omega_1} c \cos [\pi - \beta(\theta^*)] \exp (cr_v \cos [\pi - \beta(\theta^*)])dA \tag{92}$$

$$= -\int_{\omega_1} c \cos [\beta(\theta^*)] \exp (-cr_v \cos [\beta(\theta^*)])dA.$$

Hence

$$\frac{dI(r_v)}{dr_v} = c\int_{\omega_1} \cos [\beta(\theta^*)]\{\exp (cr_v \cos [\beta(\theta^*)]) - \exp (-cr_v \cos [\beta(\theta^*)])\}dA. \tag{93}$$

The right-hand side of (93) is positive. Hence Proposition II is proved.

Now let us turn back to the general problem of r variates $x^1, \cdots, x^r$ whose joint probability density function $f(x^1, \cdots, x^r, \theta^1, \cdots, \theta^k) = f(x, \theta)$ involves k unknown parameters, as considered in the previous sections.

DEFINITION II. *A sequence* $\{W_n\}$ $(n=1, \cdots, \text{ad inf.})$ *of critical regions of size* α *for testing the simple hypothesis* $\theta=\theta_0$ *is said to have asymptotically best average power with respect to the family of surfaces* K_c *and the weight function* $w(\theta)$ *if for any sequence* $\{Z_n\}$ *for which* $P(Z_n|\theta_0)=\alpha$ *we have*

$$\limsup_{n\to\infty} \left\{\text{l.u.b.}_c \left[\int_{K_c} P(Z_n|\theta)\frac{w(\theta)}{A(K_c)}dA - \int_{K_c} P(W_n|\theta)\frac{w(\theta)}{A(K_c)}dA\right]\right\} \leqq 0,$$

where

$$A(K_c) = \int_{K_c} w(\theta)dA.$$

We shall prove the following theorem.

THEOREM I. *Let* W_n^* *be a critical region for testing* $\theta=\theta_0$ *defined by the inequality*

$$n\sum_j \sum_i (\theta_n^i - \theta_0^i)(\theta_n^j - \theta_0^j)c_{ij}(\theta_n) \geqq d_n,$$

where the real number d_n *is chosen so that* $P(W_n^*|\theta_0)=\alpha$. *Denote by* S_c *the surface in the parameter space defined by the equation*

$$\sum\sum (\theta^i - \theta_0^i)(\theta^j - \theta_0^j)c_{ij}(\theta_0) = c.$$

Furthermore let $\xi(\theta)$ *be the weight function as defined in* (81) *where* $\|c_{ij}(\theta_0)\|$ *is substituted for* $\|\lambda_{ij}\|$. *Then the test* $\{W_n^*\}$ *has asymptotically best average power with respect to the family of surfaces* S_c *and the weight function* $\xi(\theta)$.

Proof. Because of Lemma 1 we can restrict ourselves to subsets of the space of the maximum likelihood estimates. Let us assume that Theorem I is not true. Then it follows from Lemma 2 that a sequence of values $\{c_n\}$ and a sequence of regions $\{Z_n^*\}$ exist such that $P(Z_n^* | \theta_0) = \alpha$ and

$$\limsup_{n=\infty} \left\{ \int_{S_{c_n}} \mathfrak{P}(Z_n^* | \theta) \zeta_n(\theta) dA - \int_{S_{c_n}} \mathfrak{P}(W_n^* | \theta) \zeta_n(\theta) dA \right\} = \delta > 0, \tag{94}$$

where

$$\zeta_n(\theta) = \xi(\theta) \Big/ \int_{S_{c_n}} \xi(\theta) dA.$$

From (94) it follows that there exists a subsequence $\{n'\}$ of the sequence $\{n\}$ such that

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \mathfrak{P}(Z_{n'}^* | \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \mathfrak{P}(W_{n'}^* | \theta) \zeta_{n'}(\theta) dA \right\} = \delta > 0. \tag{95}$$

It is easy to verify that

$$\lim_{n=\infty} \mathfrak{P}(W_{n'}^* | \theta_{n'}) = 1$$

if $\theta_{n'}$ is a point of $S_{c_{n'}}$ and if $\lim_{n=\infty} n' c_{n'} = +\infty$. Under the latter condition also

$$\lim_{n=\infty} \int_{S_{c_{n'}}} \mathfrak{P}(W_{n'}^* | \theta) \zeta_{n'}(\theta) dA = 1.$$

Thus (95) can hold only if the sequence $\{n' c_{n'}\}$ is bounded. If $\{n' c_{n'}\}$ is bounded, we obviously have for any sequence of regions $\{V_n^*\}$

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \mathfrak{P}(V_{n'}^* | \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \overline{P}(V_{n'}^* | \theta) \zeta_{n'}(\theta) dA \right\} = 0,$$

where $\overline{P}(V_n^* | \theta)$ denotes the probability of V_n^* calculated under the assumption that $n^{1/2}(\theta_n^1 - \theta^1), \cdots, n^{1/2}(\theta_n^k - \theta^k)$ have a joint normal distribution with zero means and covariance matrix equal to $\|c_{ij}(\theta_0)\|^{-1}$. Hence from (95) we obtain

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \overline{P}(Z_{n'}^* | \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \overline{P}(W_{n'}^* | \theta) \zeta_{n'}(\theta) dA \right\} = \delta > 0. \tag{96}$$

Denote by $\overline{W}_n^*$ the region defined by the inequality

$$n \sum \sum (\theta_n^i - \theta_0^i)(\theta_n^j - \theta_0^j) c_{ij}(\theta_0) \geqq d_n,$$

where d_n is chosen so that $\overline{P}(\overline{W}_n^* | \theta_0) = \alpha$. Furthermore denote by $\overline{Z}_n^*$ the sum of Z_n^* and the region $n^{1/2} | \theta_0 - \theta_n | \leqq \lambda_n$, where λ_n is chosen so that

$\overline{P}(\overline{Z}_n^* | \theta_0) = \alpha$. Since, as can easily be seen,

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \overline{P}(\overline{W}_{n'}^* | \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \overline{P}(W_{n'}^* | \theta) \zeta_{n'}(\theta) dA \right\} = 0$$

and

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \overline{P}(\overline{Z}_{n'}^* | \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \overline{P}(Z_{n'}^* | \theta) \zeta_{n'}(\theta) dA \right\} = 0,$$

we obtain from (96) a contradiction to Proposition II. Hence Theorem I is proved.

6. **Tests of simple hypotheses which have best constant power on a family of surfaces.**

DEFINITION III. *A critical region W_n for testing $\theta = \theta_0$ is said to have uniformly best constant power on the family of surfaces $\{K_c\}$ if the following two conditions are fulfilled:*

(a) $P(W_n | \theta_1) = P(W_n | \theta_2)$ *for any pair of points θ_1, θ_2 which lie on the same surface K_c.*

(b) $P(W_n | \theta) \geqq P(Z_n/\theta)$ *for any Z_n which satisfies condition* (a) *and for which $P(Z_n | \theta_0) = P(W_n | \theta_0)$.*

From Proposition II we obtain the following:

PROPOSITION III. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and a known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. Then for testing $\theta = \theta_0$, the region defined in* (82) *has uniformly best constant power on the surfaces S_c defined by the equation*

$$\sum \sum (\theta^i - \theta_0^i)(\theta^j - \theta_0^j)\lambda_{ij} = c. \tag{97}$$

Since the critical region defined in (82) satisfies condition (a) of Definition III, Proposition III is an immediate consequence of Proposition II.

DEFINITION IV. *A sequence of critical regions $\{W_n\}$ for testing $\theta = \theta_0$ is said to be of size α and to have asymptotically best constant power on the surfaces K_c if the following three conditions are fulfilled:*

(a) $P(W_n | \theta_0) = \alpha$ $(n = 1, 2, \cdots,$ ad inf.).

(b) $\lim_{n=\infty} \{\text{l.u.b.}_c [\text{l.u.b.}_{\theta \in K_c} P(W_n | \theta) - \text{g.l.b.}_{\theta \in K_c} P(W_n | \theta)]\} = 0$, *where the symbol* $\text{l.u.b.}_{\theta \in K_c}$ *means that the least upper bound is to be taken with respect to θ restricting θ to points of K_c.*

(c) *For any sequence $\{Z_n\}$ which satisfies* (a) *and* (b) *we have*

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta} [P(Z_n | \theta) - P(W_n | \theta)]\} = 0.$$

It is easy to verify that the sequence $\{W_n^*\}$ defined in Theorem I satis-

fies the conditions (a) and (b) for $K_c = S_c$, where S_c denotes the surface defined in Theorem I. Thus from Theorem I we obtain the following theorem.

THEOREM II. *Let $\{W_n^*\}$ and S_c be defined as in Theorem* I. *For testing $\theta = \theta_0$, the sequence $\{W_n^*\}$ has asymptotically best constant power on the surfaces S_c.*

7. **Most stringent tests of simple hypotheses.** Let θ and θ_0 be two parameter points and let α denote a positive number less than 1. We denote by $P_n(\theta, \theta_0, \alpha)$ the least upper bound of $P(W_n \mid \theta)$ with respect to W_n, where W_n is restricted to regions for which $P(W_n \mid \theta_0) = \alpha$. It is clear that if W_n is a critical region of size α for testing $\theta = \theta_0$, its power function can nowhere exceed the value of $P_n(\theta, \theta_0, \alpha)$, that is $P(W_n \mid \theta) \leqq P_n(\theta, \theta_0, \alpha)$ for all values of θ.

DEFINITION V. *A critical region W_n is said to be a most stringent test of the hypothesis $\theta = \theta_0$ on the level of significance α if $P(W_n \mid \theta_0) = \alpha$ and if*

$$\text{l.u.b.}_{\theta} \left[P_n(\theta, \theta_0, \alpha) - P(W_n \mid \theta) \right] \leqq \text{l.u.b.}_{\theta} \left[P_n(\theta, \theta_0, \alpha) - P(Z_n \mid \theta) \right]$$

for all regions Z_n for which $P(Z_n \mid \theta_0) = \alpha$.

We shall prove the following proposition.

PROPOSITION IV. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and a known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. Then for testing $\theta = \theta_0$ the region W defined in* (82) *is a most stringent test.*

Proof. We shall assume that Proposition IV does not hold and we shall arrive at a contradiction. If Proposition IV is not true, then there exists a region Z in the space of $y^1, \cdots, y^k$ such that $P(Z \mid \theta_0) = \alpha$ and

$$\text{(98)} \qquad \text{l.u.b.}_{\theta} \left[P(\theta, \theta_0, \alpha) - P(W \mid \theta) \right] > \text{l.u.b.}_{\theta} \left[P(\theta, \theta_0, \alpha) - P(Z \mid \theta) \right].$$

Let S_c be the surface defined by the equation

$$\sum_j \sum_i (\theta^i - \theta_0^i)(\theta^j - \theta_0^j)\lambda_{ij} = c.$$

The functions $P_1(\theta, \theta_0, \alpha)$ and $P(W \mid \theta)$ are constant on the surface S_c. Hence, on account of (98), there exists a value c_0 such that

$$P(Z \mid \theta) > P(W \mid \theta)$$

for all points θ on S_{c_0}. But this is a contradiction to Proposition II. Hence Proposition IV is proved.

DEFINITION VI. *A sequence of critical regions $\{W_n\}$ is said to be an asymptotically most stringent test of the hypothesis $\theta = \theta_0$ on the level of significance α*

if $P(W_n|\theta_0)=\alpha$ *and if for any* $\{Z_n\}$ *for which* $P(Z_n|\theta_0)=\alpha$ *we have*

$$\limsup_{n=\infty}\{\text{l.u.b.}_{\theta}[P_n(\theta,\theta_0,\alpha)-P(W_n|\theta)]-\text{l.u.b.}_{\theta}[P_n(\theta,\theta_0,\alpha)-P(Z_n|\theta)]\}\leqq 0.$$

We shall prove the following theorem.

THEOREM III. *Let* W_n^* *be the region defined in Theorem* I. *Then the sequence* $\{W_n^*\}$ *is an asymptotically most stringent test of the hypothesis* $\theta=\theta_0$.

Proof. Denote by $\mathfrak{P}_n(\theta,\theta_0,\alpha)$ the least upper bound of $\mathfrak{P}(Z_n^*|\theta)$ with respect to Z_n^*, where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which $\mathfrak{P}(Z_n^*|\theta_0)=\alpha$. Because of Lemma 1 we have

$$\lim_{n=\infty}[P_n(\theta,\theta_0,\alpha)-\mathfrak{P}_n(\theta,\theta_0,\alpha)]=0 \tag{99}$$

uniformly in θ. Denote by $\overline{P}_n(\theta,\theta_0,\alpha)$ the least upper bound of $\overline{P}(Z_n^*|\theta)$ with respect to Z_n^*, where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which $\overline{P}(Z_n^*|\theta_0)=\alpha$. The symbol $\overline{P}(V_n^*|\theta)$ denotes the probability of V_n^* calculated under the assumption that the joint distribution of $n^{1/2}(\theta_n^1-\theta^1),\cdots,n^{1/2}(\theta_n^k-\theta^k)$ is normal with zero means and covariance matrix $\|c_{ij}(\theta_0)\|^{-1}$. For any positive λ we have

$$\lim_{n=\infty}[\mathfrak{P}_n(\theta,\theta_0,\alpha)-\overline{P}_n(\theta,\theta_0,\alpha)]=0 \tag{100}$$

uniformly in θ in the domain $|n^{1/2}(\theta-\theta_0)|\leqq\lambda$. Since for any sequence $\{\theta_n\}$ for which $\lim|n^{1/2}(\theta_n-\theta_0)|=+\infty$, we have

$$\lim_{n=\infty}\mathfrak{P}_n(\theta_n,\theta_0,\alpha)=\lim_{n=\infty}\overline{P}_n(\theta_n,\theta_0,\alpha)=1,$$

we obtain from (100)

$$\lim_{n=\infty}[\mathfrak{P}_n(\theta,\theta_0,\alpha)-\overline{P}_n(\theta,\theta_0,\alpha)]=0 \tag{101}$$

uniformly in θ. For any c let S_c be the surface defined by

$$\sum_{j=1}^{k}\sum_{i=1}^{k}(\theta^i-\theta_0^i)(\theta^j-\theta_0^j)c_{ij}(\theta_0)=c.$$

Obviously $\overline{P}_n(\theta,\theta_0,\alpha)$ is constant along the surface S_c. From (99) and (101) we obtain

$$\lim_{n=\infty}\{\text{l.u.b.}_{\theta\in S_c}P_n(\theta,\theta_0,\alpha)-\text{g.l.b.}_{\theta\in S_c}P_n(\theta,\theta_0,\alpha)\}=0 \tag{102}$$

uniformly in θ. We shall derive a contradiction from the assumption that Theorem III is not true. If Theorem III is not true, there exists a sequence $\{Z_n\}$ of regions such that $P(Z_n|\theta_0)=\alpha$ and

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta} [P_n(\theta, \theta_0, \alpha) - P(W_n^* | \theta)] - \text{l.u.b.}_{\theta} [P_n(\theta, \theta_0, \alpha) - P(Z_n | \theta)]\} = \delta > 0. \tag{103}$$

On account of (102) and since

$$\lim_{n=\infty} [\text{l.u.b.}_{\theta \in S_c} P(W_n^* | \theta) - \text{g.l.b.}_{\theta \in S_c} P(W_n^* | \theta)] = 0$$

uniformly in c (see Theorem II), we obtain from (103) that there exists a sequence $\{c_n\}$ and a subsequence $\{n'\}$ of $\{n\}$ such that for all points $\theta_{n'}$ of $S_{c_{n'}}$

$$P(Z_{n'} | \theta_{n'}) > P(W_{n'}^* | \theta_{n'}) + \delta/2$$

for all n greater than a certain n_0. But this contradicts Theorem I. Hence Theorem III is proved.

8. **Definitions of "best" tests of composite hypotheses.** In this section we shall extend the definitions given in the previous sections to the case of composite hypotheses. Let ω be a subset of the parameter space and denote by H_ω the hypothesis that the true parameter point is contained in ω. In all that follows the letter θ printed in boldface will indicate that the parameter point lies in ω. For example, the symbol $\text{l.u.b.}_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$ denotes the least upper bound of the function $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ where $\boldsymbol{\theta}$ is restricted to points of ω. For any point $\boldsymbol{\theta}$ and for any real value c let $K_c(\boldsymbol{\theta})$ denote a surface in the parameter space. For instance $K_c(\boldsymbol{\theta})$ may be given by r equations in θ

$$\phi_1(\theta, \boldsymbol{\theta}) = \cdots = \phi_r(\theta, \boldsymbol{\theta}) = 0,$$

where $\phi_1(\theta, \boldsymbol{\theta}), \cdots, \phi_r(\theta, \boldsymbol{\theta})$ are some analytic functions of θ and $\boldsymbol{\theta}$.

DEFINITION VII. *A critical region W_n for testing H_ω is said to have uniformly best average power with respect to a family of surfaces $K_c(\boldsymbol{\theta})$ and a weight function $w(\theta)$ if for any Z_n for which*

$$\text{l.u.b.}_{\boldsymbol{\theta}} P(Z_n | \boldsymbol{\theta}) = \text{l.u.b.}_{\boldsymbol{\theta}} P(W_n | \boldsymbol{\theta})$$

we have

$$\int_{K_c(\boldsymbol{\theta})} P(W_n | \theta) w(\theta) dA \geqq \int_{K_c(\boldsymbol{\theta})} P(Z_n | \theta) w(\theta) dA$$

for any $\boldsymbol{\theta}$ and for any c for which $K_c(\boldsymbol{\theta})$ is defined.

DEFINITION VIII. *A sequence $\{W_n\}$ ($n = 1, 2, \cdots$, ad inf.) of critical regions for testing the hypothesis H_ω is said to have asymptotically best average power with respect to a family of surfaces $K_c(\boldsymbol{\theta})$ and a weight function $w(\theta)$ if the following two conditions are fulfilled:*

(a) *There exists a fixed* α *such that*

$$\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha \qquad (n = 1, 2, \cdots, \text{ad inf.})$$

(b) *For any sequence* $\{Z_n\}$ *for which* $\text{l.u.b.}_{\theta}\, P(Z_n \mid \theta) = \alpha$, *we have*

$$\limsup_{n=\infty} \left\{ \text{l.u.b.}_{c,\theta} \left[\int_{K_c(\theta)} P(Z_n \mid \theta) \frac{w(\theta)}{A[K_c(\theta)]}\, dA - \int_{K_c(\theta)} P(W_n \mid \theta) \frac{w(\theta)}{A[K_c(\theta)]}\, dA \right] \right\} \leqq 0,$$

where

$$A[K_c(\theta)] = \int_{K_c(\theta)} w(\theta)\, dA.$$

DEFINITION IX. *A critical region* W_n *for testing* H_ω *is said to have uniformly best constant power on the family of surfaces* $K_c(\theta)$ *if the following two conditions are fulfilled*:

(a) $P(W_n \mid \theta') = P(W_n \mid \theta'')$ *for all pairs of points* θ' *and* θ'' *which lie on the same surface* $K_c(\theta)$.

(b) $P(W_n \mid \theta) \geqq P(Z_n \mid \theta)$ *for any* θ *not in* ω *and for any* Z_n *which satisfies* (a) *and the condition*

$$\text{l.u.b.}_{\theta}\, P(Z_n \mid \theta) = \text{l.u.b.}_{\theta}\, P(W_n \mid \theta).$$

DEFINITION X. *A sequence of critical regions* $\{W_n\}$ *for testing* H_ω *is said to have asymptotically best constant power on the surfaces* $K_c(\theta)$ *if the following three conditions are fulfilled*:

(a) $\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha$ $(n = 1, 2, \cdots, \text{ad inf.})$.

(b) $\lim_{n=\infty} \{\text{l.u.b.}_{c,\theta} [\text{l.u.b.}_{\theta \in K_c(\theta)} P(W_n \mid \theta) - \text{g.l.b.}_{\theta \in K_c(\theta)} P(W_n \mid \theta)]\} = 0$.

(c) *For any sequence* $\{Z_n\}$ *which satisfies* (a) *and* (b) *we have*

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta \in \bar{\omega}} [P(Z_n \mid \theta) - P(W_n \mid \theta)]\} = 0,$$

where $\bar{\omega}$ *is the complement of* ω.

DEFINITION XI. *Denote by* $P_n(\theta, \omega, \alpha)$ *the least upper bound of* $P(Z_n \mid \theta)$ *with respect to* Z_n *subject to the condition* $\text{l.u.b.}_{\theta}\, P(Z_n \mid \theta) = \alpha$. *A critical region* W_n *is said to be a most stringent test of the hypothesis* H_ω *if for some positive* α

$$\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha$$

and

$$\text{l.u.b.}_{\theta} [P_n(\theta, \omega, \alpha) - P(W_n \mid \theta)] \leqq \text{l.u.b.}_{\theta} [P_n(\theta, \omega, \alpha) - P(Z_n \mid \theta)]$$

for all regions Z_n *for which* $\text{l.u.b.}_{\theta}\, P(Z_n \mid \theta) = \alpha$.

DEFINITION XII. *A sequence of critical regions* $\{W_n\}$ *is said to be an asymptotically most stringent test of the hypothesis* H_ω *if the following two conditions are fulfilled*:

(a) *There exists a positive* α *such that*

$$\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha \qquad (n = 1, 2, \cdots, \text{ad inf.}).$$

(b) *For any sequence* $\{Z_n\}$ *which satisfies* (a) *for the same* α *we have*

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta}\, [P_n(\theta, \omega, \alpha) - P(W_n \mid \theta)] - \text{l.u.b.}_{\theta}\, [P_n(\theta, \omega, \alpha) - P(Z_n \mid \theta)]\} \leqq 0.$$

In Definitions VII–XII we have formulated the condition

$$\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha.$$

The question can be raised whether, in place of this condition, the requirement that

$$(*) \qquad P(W_n \mid \theta) = \alpha$$

for all points θ should be made; or whether the weaker condition that

$$(**) \qquad \lim_{n=\infty} P(W_n \mid \theta) = \alpha$$

uniformly in θ should be required. Condition (*) has the serious drawback that regions satisfying it do not always exist. Even in cases where (*) can be fulfilled, it imposes too strong a restriction on the possible choice of W_n, which does not seem to be quite justified. It is conceivable that in some cases a region W_n' may exist which does not satisfy (*) but has such an advantageous power function that we prefer it to any region W_n which satisfies(*).

As to the condition (**), we shall see that it is satisfied for the sequence $\{W_n\}$ which is shown in this paper to be asymptotically best according to all three definitions VIII, X and XII. Hence the same sequence $\{W_n\}$ remains asymptotically best if we replace the condition l.u.b.$_\theta$ $P(W_n \mid \theta) = \alpha$ by (**) in the definitions VIII, X and XII.

In the following §§9–11 we shall discuss a linear hypothesis of the following type: $\theta^1 = \theta_0^1, \cdots, \theta^r = \theta_0^r$ $(r < k)$, where $\theta_0^1, \cdots, \theta_0^r$ are some specified values. That is to say, the set ω is the set of all points θ for which the above equations hold. In §12 the general composite hypothesis will be discussed.

9. Tests of linear composite hypotheses which have uniformly best average power over a family of surfaces. Let H_ω be the hypothesis that $\theta^1 = \theta_0^1, \cdots, \theta^r = \theta_0^r$ $(r < k)$. We shall introduce the following notation: For any parameter point $\theta = (\theta^1, \cdots, \theta^k)$ the symbol ${}_1\theta$ will denote the vector in the r-dimensional space with the components $\theta^1, \cdots, \theta^r$, and ${}_2\theta$ will denote the vector in the $k-r$ dimensional space with the components $\theta^{r+1}, \cdots, \theta^k$. For any function $\psi(\theta)$ of θ we shall use the synonymous nota-

tion $\psi({}_1\theta, {}_2\theta)$. For instance $P(W|{}_1\theta, {}_2\theta)$ is synonymous with $P(W|\theta)$.

Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$ $(i, j = 1, \cdots, k)$, which is nonsingular. Denote by W the region in the space of $y^1, \cdots, y^k$ given by the inequality

$$\sum_{q=1}^{r}\sum_{p=1}^{r}\lambda_{pq}(y^p - \theta_0^p)(y^q - \theta_0^q) \geqq d, \tag{104}$$

where $\|\lambda_{pq}\| = \|\sigma_{pq}\|^{-1}$ $(p, q = 1, \cdots, r)$. Consider the nonsingular linear transformation of the variates $y^1, \cdots, y^k$ given by the equations

$$\begin{aligned} y'^p - \theta_0^p &= \beta_{p1}(y^1 - \theta_0^1) + \cdots + \beta_{pr}(y^r - \theta_0^r) && (p = 1, \cdots, r), \\ y'^t &= \gamma_{t1}y^1 + \cdots + \gamma_{tk}y^k && (t = r+1, \cdots, k), \end{aligned} \tag{105}$$

such that $y'^1, \cdots, y'^k$ are independently distributed with unit variances. Denote by $S_c(\theta)$ the surface given by the equations

$$\begin{aligned} &\sum_{q=1}^{r}\sum_{p=1}^{r}\lambda_{pq}(\theta^p - \theta_0^p)(\theta^q - \theta_0^q) = c, \\ &\gamma_{t1}\theta^1 + \cdots + \gamma_{tk}\theta^k = \sum_{i=1}^{k}\gamma_{ti}\theta_0^i \qquad (t = r+1, \cdots, k). \end{aligned} \tag{106}$$

Consider the transformation of the parameter space given by

$$\begin{aligned} \theta'^p - \theta_0^p &= \beta_{p1}(\theta^1 - \theta_0^1) + \cdots + \beta_{pr}(\theta^r - \theta_0^r) && (p = 1, \cdots, r), \\ \theta'^t &= \gamma_{t1}\theta^1 + \cdots + \gamma_{tk}\theta^k && (t = r+1, \cdots, k), \end{aligned} \tag{107}$$

where the coefficients β_{pq} and γ_{ti} are the same as in (105). The transformation (107) transforms the surface $S_c(\theta)$ into a sphere $S'_c(\theta)$ given by

$$\sum_{p=1}^{r}(\theta'^p - \theta_0^p)^2 = c, \qquad \theta'^t = \sum_{i=1}^{k}\gamma_{ti}\theta_0^i = \theta_0'^t. \tag{108}$$

For any point θ and for any positive ρ consider the set $\omega(\theta, \rho)$ consisting of all points ${}^*\theta$ which lie on the same $S_c(\theta)$ as θ, and for which $|{}^*\theta - \theta| \leqq \rho$. Let

$$\xi(\theta) = \lim_{\rho=0}\frac{A[\omega'(\theta, \rho)]}{A[\omega(\theta, \rho)]}, \tag{109}$$

where $\omega'(\theta, \rho)$ is the image of $\omega(\theta, \rho)$ (by transformation (107)) and, for any set ω, $A(\omega)$ denotes the $(r-1)$-dimensional area of ω. We shall prove the following proposition.

PROPOSITION V. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. For testing the hypothesis ${}_1\theta = {}_1\theta_0$ on the basis of a single*

observation on each of the variates $y^1, \cdots, y^k$, *the critical region* W *given in* (104) *has uniformly best average power with respect to the family of surfaces* $S_c(\theta)$ *defined in* (106) *and the weight function* $\xi(\theta)$ *given in* (109).

Proof. Because of (109) we have

$$\int_{S_c(\theta)} P(Z \mid \theta)\xi(\theta)dA = \int_{S_c'(\theta)} P(Z' \mid \theta')dA, \tag{110}$$

where Z denotes an arbitrary region in the space of $y^1, \cdots, y^k$ and Z' is the image of Z by transformation (105). The region W is transformed into W' given by

$$(y'^1 - \theta_0^1)^2 + \cdots + (y'^r - \theta_0^r)^2 \geqq d. \tag{111}$$

In order to prove Proposition V we have merely to show that

$$\int_{S_c'(\theta)} P(W' \mid \theta')dA \geqq \int_{S_c'(\theta)} P(Z' \mid \theta')dA \tag{112}$$

for any $c>0$, for any θ, and for any region Z' in the space of $y'^1, \cdots, y'^k$ for which l.u.b.$_{{}_2\theta'}$ $P(Z' \mid {}_1\theta_0, {}_2\theta') = $ l.u.b.$_{{}_2\theta'}$ $P(W' \mid {}_1\theta_0, {}_2\theta')$. For any point θ' of $S_c'(\theta)$ we have ${}_2\theta' = {}_2\theta'$. By a lemma of Neyman and Pearson, (112) is proved if we can show that there exists a function $d(c)$ such that

$$\left.\begin{aligned} \frac{\int_{S_c'(\theta)} p(y'^1, \cdots, y'^k \mid {}_1\theta', {}_2\theta')dA}{p(y'^1, \cdots, y'^k \mid {}_1\theta_0, {}_2\theta')} &\geqq d(c) \text{ within } W', \\ &\leqq d(c) \text{ outside } W' \end{aligned}\right\} \tag{113}$$

for all values of c and θ where $p(y'^1, \cdots, y'^k \mid \theta')$ denotes the joint probability density of $y'^1, \cdots, y'^k$ under the assumption that θ' is the true parameter point. Obviously

$$\frac{p(y'^1, \cdots, y'^k \mid {}_1\theta', {}_2\theta')}{p(y'^1, \cdots, y'^k \mid {}_1\theta_0, {}_2\theta')} = \frac{p(y'^1, \cdots, y'^r \mid {}_1\theta')}{p(y'^1, \cdots, y'^r \mid {}_1\theta_0)}.$$

Hence (113) is equivalent to

$$\left.\begin{aligned} \frac{\int_{S_c'(\theta)} p(y'^1, \cdots, y'^r \mid {}_1\theta')dA}{p(y'^1, \cdots, y'^r \mid {}_1\theta_0)} &\geqq d(c) \text{ within } W', \\ &\leqq d(c) \text{ outside } W'. \end{aligned}\right\} \tag{114}$$

The proof of (114) is omitted, since it is the same as that of the inequalities (87) and (88). Hence Proposition V is proved.

Let the critical region W_n^* be defined by the following inequality

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}(\theta_n^p - \theta_0^p)(\theta_n^q - \theta_0^q)\bar{c}_{pq}(\theta_n) \geqq d_n, \tag{115}$$

where $\|\bar{c}_{pq}(\theta)\| = \|\sigma_{pq}(\theta)\|^{-1}$ $(p, q = 1, \cdots, r)$ and $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$ $(i, j$

$=1, \cdots, k$). The constant d_n is chosen so that

$$\text{l.u.b.}_{\theta} P(W^* \mid \theta) = \alpha.$$

Let $Z_n^i(\theta) = n^{1/2}(\theta_n^i - \theta^i)$ $(i=1, \cdots, k)$ and consider the nonsingular linear transformation

$$\begin{aligned} \overline{Z}_n^p(\theta) &= \beta_{p1}(\theta)Z_n^1(\theta) + \cdots + \beta_{pr}(\theta)Z_n^r(\theta) \qquad (p = 1, \cdots, r), \\ \overline{Z}_n^t(\theta) &= \gamma_{t1}(\theta)Z_n^1(\theta) + \cdots + \gamma_{tk}(\theta)Z_n^k(\theta) \qquad (t = r+1, \cdots, k), \end{aligned} \tag{116}$$

such that $\overline{Z}_n^1(\theta), \cdots, \overline{Z}_n^k(\theta)$ would be independently distributed with unit variances if the covariance matrix of $Z_n^1(\theta), \cdots, Z_n^k(\theta)$ were given by $\|\sigma_{ij}(\theta)\|$. Denote by $S_c(\theta)$ the surface defined by the equations

$$\begin{aligned} &\sum_{q=1}^{r}\sum_{p=1}^{r}(\theta^p - \theta_0^p)(\theta^q - \theta_0^q)\bar{c}_{pq}(\theta) = c, \\ &\gamma_{t1}(\theta)\theta^1 + \cdots + \gamma_{tk}(\theta)\theta^k = \sum \gamma_{ti}(\theta)\theta^i \quad (t = r+1, \cdots, k). \end{aligned} \tag{117}$$

For any positive δ denote by S_δ the set of all points θ for which $|\theta^p - \theta_0^p| \leqq \delta$ $(p=1, \cdots, r)$. We shall prove the existence of a positive δ such that for any point θ in S_δ there exists exactly one surface $S_c(\theta)$, that is exactly one value of c and exactly one point θ, such that θ lies on the surface $S_c(\theta)$. This statement is obviously proved if we show that for any point θ in S_δ the set of $k-r$ equations

$$\gamma_{t1}(\theta)(\theta^1 - \theta^1) + \cdots + \gamma_{tk}(\theta)(\theta^k - \theta^k) = 0 \qquad (t = r+1, \cdots, k) \tag{118}$$

has a unique solution in the unknowns $\theta^{r+1}, \cdots, \theta^k$. From the definition of the quantities $\beta_{pq}(\theta)$ and $\gamma_{ti}(\theta)$ it follows that

$$A(\theta)\|\sigma_{ij}(\theta)\|\overline{A}(\theta) = I, \tag{119}$$

where $A(\theta)$ denotes the matrix

$$\left\| \begin{array}{ccccccc} \beta_{11}(\theta) & \beta_{12}(\theta) & \cdots & \beta_{1r}(\theta) & 0 & \cdots & 0 \\ \beta_{12}(\theta) & \beta_{22}(\theta) & \cdots & \beta_{2r}(\theta) & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \beta_{r1}(\theta) & \beta_{r2}(\theta) & \cdots & \beta_{rr}(\theta) & 0 & \cdots & 0 \\ \gamma_{r+1\,1}(\theta) & \gamma_{r+1\,2}(\theta) & \cdots & \gamma_{r+1\,r}(\theta) & \gamma_{r+1\,r+1}(\theta) & \cdots & \gamma_{r+1\,k}(\theta) \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \gamma_{k1}(\theta) & \gamma_{k2}(\theta) & \cdots & \gamma_{kr}(\theta) & \gamma_{k\,r+1}(\theta) & \cdots & \gamma_{kk}(\theta) \end{array} \right\|, \tag{120}$$

$\overline{A}(\theta)$ is the transposed of $A(\theta)$ and I denotes the unit matrix. Since $\sigma_{ij}(\theta)$ is a continuous and bounded function of θ and since the determinant $|\sigma_{ij}(\theta)|$ has a positive lower bound, we find that $\beta_{pq}(\theta)$ and $\gamma_{ti}(\theta)$ are continuous and

bounded functions of θ and that the absolute value of the determinant $A(\theta)$ has a positive lower bound. Hence also the absolute value of the determinant

$$\gamma(\theta) = \begin{vmatrix} \gamma_{r+1\,r+1}(\theta) & \cdots & \gamma_{r+1\,k}(\theta) \\ \cdot & \cdots & \cdot \\ \gamma_{k\,r+1}(\theta) & \cdots & \gamma_{kk}(\theta) \end{vmatrix} \tag{121}$$

has a positive lower bound.

Let $\theta = \theta^*$ where θ^* denotes an arbitrary point of ω. Then, since the determinant in (121) has a positive lower bound, the equations (118) have a unique solution in the unknowns $\theta^{r+1}, \cdots, \theta^k$, namely the solution $\theta = \theta^*$. Furthermore we see that the Jacobian of the equations (118), taken at the point $\theta = \theta^*$, is equal to $\gamma(\theta^*)$. Since the absolute value of $\gamma(\theta^*)$ has a positive lower bound, there exists a positive δ such that the equations (118) have a unique solution in θ if $|\theta - \theta^*| \leqq \delta$. This proves the existence of a positive δ such that for any point θ in S_δ there exists exactly one surface $S_c(\theta)$ such that θ lies on $S_c(\theta)$.

Since for the critical region W_n^* defined in (115) we obviously have $\lim_{n=\infty} P(W_n^* | \theta) = 1$ uniformly over the domain $|\theta^p - \theta_0^p| \geqq \delta$, we shall restrict ourselves to the consideration of points θ for which $|\theta^p - \theta_0^p| \leqq \delta$ $(p = 1, \cdots, r)$.

Consider the transformation of the parameter space given by

$$\begin{aligned} \theta'^p = \theta_0^p = \beta_{p1}(\theta)(\theta^1 - \theta_0^1) + \cdots + \beta_{pr}(\theta)(\theta^r - \theta_0^r) \qquad & (p = 1, \cdots, r), \\ \theta'^t = \gamma_{t1}(\theta)\theta^1 + \cdots + \gamma_{tk}(\theta)\theta^k \qquad & (t = r+1, \cdots, k), \end{aligned} \tag{122}$$

where θ denotes the point for which θ lies on $S_c(\theta)$. The transformation (122) transforms $S_c(\theta)$ into the sphere $S_c'(\theta)$ given by

$$\sum_{p=1}^{r} (\theta'^p - \theta_0^p)^2 = c; \quad \theta'^t = \sum \gamma_{ti}(\theta)\theta^i = \theta'^t \qquad (t = r+1, \cdots, k). \tag{123}$$

We define a weight function $\xi(\theta)$ as follows:

$$\xi(\theta) = \lim_{\rho=0} \frac{A[\omega'(\theta, \rho)]}{A[\omega(\theta, \rho)]}, \tag{124}$$

where the symbols on the right-hand side have the same meaning as in (109).

THEOREM IV. *Let the critical region W_n^* for testing ${}_1\theta = {}_1\theta_0$ be the region defined in* (115). *Furthermore, let $S_c(\theta)$ be the surface defined in* (117) *and let $\xi(\theta)$ be the weight function defined in* (124). *Then $\{W_n^*\}$ has asymptotically best average power with respect to the family of surfaces $S_c(\theta)$ and the weight function $\xi(\theta)$.*

Proof. Because of Lemma 1, we can restrict ourselves to subsets of the space of the maximum likelihood estimates. Because of Lemma 2, Theo-

rem IV is proved if we show that for any $\{Z_n^*\}$ for which l.u.b.$_\theta$ $P(Z_n^*|\theta)=\alpha$ we have

$$\limsup_{n=\infty}\left\{\text{l.u.b.}_{c,\theta}\left[\int_{S_c(\theta)}\mathfrak{P}(Z_n^*|\theta)\zeta_n(\theta)dA-\int_{S_c(\theta)}\mathfrak{P}(W_n^*|\theta)\zeta_n(\theta)dA\right]\right\}\leqq 0,$$

where

$$\zeta_n(\theta)=\xi(\theta)\Big/\int_{S_c(\theta)}\xi(\theta)dA.$$

If Theorem IV were not true, there would exist a sequence $\{c_n\}$, a sequence $\{\theta_n\}$, a sequence $\{Z_n^*\}$, and a subsequence $\{n'\}$ of $\{n\}$ such that

$$\text{l.u.b.}_{\theta}\ P(Z_n^*|\theta)=\alpha,$$

$$\lim_{n=\infty}\left\{\int_{S_{c_{n'}}(\theta_{n'})}\mathfrak{P}(Z_{n'}^*|\theta)\zeta_{n'}(\theta)dA-\int_{S_{c_{n'}}(\theta_{n'})}\mathfrak{P}(W_{n'}^*|\theta)\zeta_{n'}(\theta)dA\right\}=\delta>0. \tag{125}$$

It is easy to verify that for any sequence $\{c_n\}$ for which $\lim nc_n=+\infty$ we have $\int_{S_{c_n}(\theta)}\mathfrak{P}(W_n^*|\theta)\zeta_n(\theta)dA=1$ uniformly in θ. Hence (125) can hold only if the sequence $\{n'c_{n'}\}$ is bounded. If $\{n'c_{n'}\}$ is bounded, for any sequence of regions $\{V_n^*\}$ we obviously have

$$\lim_{n=\infty}\left\{\int_{S_{c_{n'}}(\theta_{n'})}\mathfrak{P}(V_{n'}^*|\theta)\zeta_{n'}(\theta)dA-\int_{S_{c_{n'}}(\theta_{n'})}P_{\theta_{n'}}(V_{n'}^*|\theta)\zeta_{n'}(\theta)dA\right\}=0, \tag{126}$$

where $P_\theta(V_n^*|\theta)$ denotes the probability of V_n^* calculated under the assumption that $n^{1/2}(\theta_n^1-\theta^1),\cdots,n^{1/2}(\theta_n^k-\theta^k)$ have a joint normal distribution with zero means and covariance matrix $\sigma_{ij}(\theta)$. Let $W_n^*(\theta)$ be the region defined by

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}(\theta_n^p-\theta_0^p)(\theta_n^q-\theta_0^q)\bar{c}_{pq}(\theta)\geqq d_n.$$

It is clear that if $\{n'c_{n'}\}$ is bounded, we have

$$\lim_{n=\infty}\{P_\theta(W_{n'}^*|\theta)-P_\theta[W_{n'}^*(\theta)|\theta]\}=0 \tag{127}$$

uniformly in θ and θ over the domain in which θ is a point of $S_{c_{n'}}(\theta)$. From (125), (126) and (127) we obtain

$$\lim_{n=\infty}\left\{\int_{S_{c_{n'}}(\theta_{n'})}P_{\theta_{n'}}(Z_{n'}^*|\theta)\zeta_{n'}(\theta)dA-\int_{S_{c_{n'}}(\theta_{n'})}P_{\theta_{n'}}[W_{n'}^*(\theta_{n'})|\theta]\zeta_{n'}(\theta)dA\right\}=\delta>0. \tag{128}$$

The surface $S_c(\theta)$ defined in (106) is identical with the surface $S_c(\theta)$ defined

in (117) if in (106) we substitute $c_{ij}(\theta)$ for λ_{ij}, that is we substitute $\bar{c}_{pq}(\theta)$ for $\bar{\lambda}_{pq}$. Similarly, if we substitute $c_{ij}(\theta)$ for λ_{ij} for any point of the surface $S_c(\theta)$, the value of the weight function $\xi(\theta)$ defined in (109) is the same as the value of $\xi(\theta)$ defined in (124). Hence, since

$$\lim_{n=\infty} [\text{l.u.b.}_{\theta} P_{\theta}(Z_n^* \mid \theta)] = \alpha,$$

equation (128) is in contradiction to Proposition V. Thus Theorem IV is proved.

10. **Tests of linear composite hypotheses which have best constant power on a family of surfaces.** The critical region W defined in (104) satisfies condition (a) of Definition IX if $K_c(\theta)$ is equal to $S_c(\theta)$ given in (106). Hence from Proposition V we obtain the following proposition.

PROPOSITION VI. *The region W given in* (104) *has uniformly best constant power on the surfaces $S_c(\theta)$ defined in* (106).

If W_n^* is the region defined in (115) and if $K_c(\theta)$ is equal to the surface $S_c(\theta)$ defined in (117), then $\{W_n^*\}$ satisfies conditions (a) and (b) of Definition X. Hence, from Theorem IV we easily obtain the following theorem.

THEOREM V. *Let W_n^* be the region defined in* (115) *and let $S_c(\theta)$ be the surface defined in* (117), *then for testing ${}_1\theta = {}_1\theta_0$, $\{W_n^*\}$ has asymptotically best constant power along the surfaces $S_c(\theta)$.*

11. **Most stringent tests of linear composite hypotheses.** We shall prove the following proposition.

PROPOSITION VII. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. For testing the hypothesis ${}_1\theta = {}_1\theta_0$ on the basis of a single observation on each of the variates $y^1, \cdots, y^k$, the region W given in* (104) *is a most stringent test.*

Proof. First we shall show that $P(\theta, \omega, \alpha)$ is constant along $S_c(\theta)$ where $S_c(\theta)$ is defined in (106). Consider a linear transformation of $y^1, \cdots, y^k$ as defined in (105). Then the transformed variates $y'^1, \cdots, y'^k$ are independently distributed with unit variances. Denote by θ' the image of θ obtained by the transformation (107) and let $P'(\theta', \omega, \alpha)$ be equal to l.u.b. $P(Z' \mid \theta')$ with respect to Z', where Z' may be any region in the space of $y'^1, \cdots, y'^k$ subject to the condition that

$$\text{l.u.b.}_{{}_2\theta'} P(Z' \mid {}_1\theta_0, {}_2\theta') = \alpha.$$

Obviously $P'(\theta', \omega, \alpha) = P(\theta, \omega, \alpha)$.

Hence we have merely to show that $P'(\theta', \omega, \alpha)$ is constant along $S_c'(\theta)$ where $S_c'(\theta)$ is the image of $S_c(\theta)$ and is given by

$$\sum_{p=1}^{r} (\theta'^p - \theta_0^p)^2 = c; \qquad \theta'^t = \theta'^t \qquad (t = r+1, \cdots, k).$$

Let $P^*(\theta', {}_1\theta_0, \alpha)$ be equal to the least upper bound of $P(Z'|\theta')$ with respect to all regions Z' in the space of $y'^1, \cdots, y'^k$ for which $P(Z'|{}_1\theta_0, {}_2\theta') = \alpha$. Obviously $P^*(\theta', {}_1\theta_0, \alpha) \geqq P'(\theta', \omega, \alpha)$. It is easy to verify that the region V' for which $P(V'|{}_1\theta', {}_2\theta') = P^*(\theta', {}_1\theta_0, \alpha)$ is a subset in the space of $y'^1, \cdots, y'^r$. Hence

$$P(V'|\,{}_1\theta_0, {}_2\theta_1') = P(V'|\,{}_1\theta_0, {}_2\theta_2')$$

for any pair of points θ_1' and θ_2', and therefore

$$P^*(\theta', {}_1\theta_0, \alpha) = P'(\theta', \omega, \alpha) = P(\theta, \omega, \alpha).$$

Since $P^*(\theta', {}_1\theta_0, \alpha)$ is constant along $S_c(\theta)$, our statement is proved. From this and Proposition V, Proposition VII easily follows.

THEOREM VI. *Let W_n^* be the region defined in* (115). *Then $\{W_n^*\}$ is an asymptotically most stringent test of the hypothesis ${}_1\theta = {}_1\theta_0$.*

Proof. Denote by $\mathfrak{P}_n(\theta, \omega, \alpha)$ the least upper bound of $\mathfrak{P}(Z_n^*|\theta)$ with respect to Z_n^*, where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which

$$\text{l.u.b.}_{\theta}\, \mathfrak{P}(Z_n^*|\,\theta) = \alpha.$$

On account of Lemmas 1 and 2 we have

$$\lim_{n=\infty} \{P_n(\theta, \omega, \alpha) - \mathfrak{P}_n(\theta, \omega, \alpha)\} = 0 \tag{129}$$

uniformly in θ.

Denote by $\overline{P}_n(\theta, {}_1\theta_0, \alpha)$ the least upper bound of $\overline{P}_n(Z_n^*|\theta)$ with respect to Z_n^* where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which

$$\text{l.u.b.}_{{}_2\theta}\, \overline{P}(Z_n^*|\,{}_1\theta_0, {}_2\theta) = \alpha.$$

The symbol $\overline{P}(V_n^*|\theta)$ denotes the probability of V_n^* calculated under the assumption that the joint distribution of $n^{1/2}(\theta_n^1 - \theta^1), \cdots, n^{1/2}(\theta_n^k - \theta^k)$ is normal with zero means and covariance matrix $\|c_{ij}(\theta)\|^{-1}$, where θ denotes that point for which θ lies on the surface $S_c(\theta)$ defined in (117). It can be shown that for any positive λ we have

$$\lim_{n=\infty} \{\mathfrak{P}_n(\theta, \omega, \alpha) - \overline{P}_n(\theta, {}_1\theta_0, \alpha)\} = 0 \tag{130}$$

uniformly in θ in the domain $|{}_1\theta - {}_1\theta_0| \leqq \lambda/n^{1/2}$. Since for any sequence $\{\theta_n\}$ for which

$$\lim_{n=\infty} n^{1/2}\left(\sum_{p=1}^{r}(\theta^p - \theta_0^p)^2\right)^{1/2} = +\infty,$$

we have

$$\lim_{n=\infty} \mathfrak{P}(\theta_n, \omega, \alpha) = \lim_{n=\infty} \overline{P}_n(\theta_n, {}_1\theta_0, \alpha) = 1,$$

we obtain from (130)

$$\lim_{n=\infty} \{\mathfrak{P}_n(\theta, \omega, \alpha) - \overline{P}_n(\theta, {}_1\theta_0, \alpha)\} = 0 \tag{131}$$

uniformly in θ. The function $\overline{P}_n(\theta, {}_1\theta_0, \alpha)$ is constant along the surface $S_c(\theta)$ defined in (117). This can be proved in the same way as the constancy of $P(\theta, \omega, \alpha)$ on $S_c(\theta)$ defined in (106). Hence from (129) and (131) we obtain

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta \in S_c(\omega)} P_n(\theta, \omega, \alpha) - \text{g.l.b.}_{\theta \in S_c(\omega)} P_n(\theta, \omega, \alpha)\} = 0 \tag{132}$$

uniformly in θ and c. According to Theorem V we have

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta \in S_c(\omega)} P(W_n^* | \theta) - \text{g.l.b.}_{\theta \in S_c(\omega)} P(W_n^* | \theta)\} = 0 \tag{133}$$

uniformly in c and θ. Theorem VI follows from (132), (133) and Theorem IV.

12. **The general composite hypothesis.** In §§9–11 we have considered the linear composite hypothesis ${}_1\theta = {}_1\theta_0$. Now we shall discuss a general composite hypothesis H_ω where ω denotes a subset of the parameter space given by r equations

$$\xi^1(\theta) = \xi^2(\theta) = \cdots = \xi^r(\theta) = 0 \qquad (r < k), \tag{134}$$

that is, ω is the set of all points θ which satisfy equations (134). We make the following assumption.

ASSUMPTION VI. *There exist $k-r$ functions $\xi^{r+1}(\theta), \cdots, \xi^k(\theta)$ such that the following three conditions are fulfilled:*

(a) *The transformation which transforms the point θ into the point ξ with the coordinates $\xi^1(\theta), \cdots, \xi^k(\theta)$ is a topological transformation of Ω into itself.*

(b) *The first and second order partial derivatives of $\xi^1(\theta), \cdots, \xi^k(\theta)$ are uniformly continuous and bounded functions of θ.*

(c) *The greatest lower bound of the absolute value of the Jacobian $\partial(\xi^1, \cdots, \xi^k)/\partial(\theta^1, \cdots, \theta^k)$ is positive.*

Let $\xi = (\xi^1, \cdots, \xi^k)$ denote a variable point of the parameter space Ω.

Since according to Assumption VI the transformation

$$\xi^i = \xi^i(\theta) \qquad (i = 1, \cdots, k) \tag{135}$$

is topological, we can solve the equations (135) and we obtain

$$\theta^i = \theta^i(\xi^1, \cdots, \xi^k) \qquad (i = 1, \cdots, k). \tag{136}$$

From conditions (b) and (c) of Assumption VI it follows that the first and second order derivatives of $\theta^i(\xi^1, \cdots, \xi^k)$ are uniformly continuous and bounded functions of ξ and the absolute value of the Jacobian $\partial(\theta^1, \cdots, \theta^k)/\partial(\xi^1, \cdots, \xi^k)$ has a positive lower bound.

Let $f^*(x, \xi)$ be the probability density function we obtain from the probability density function $f(x, \theta)$ of x by substituting the right-hand side of (136) for θ^i. Hence $f^*(x, \xi)$ is the probability density function of x in the transformed parameter space. It is clear that the maximum likelihood estimate of ξ^i is equal to $\hat{\xi}_n^i = \xi^i(\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k)$, where $\hat{\theta}_n$ is the maximum likelihood estimate of θ.

Denote by I*, II*, $\cdots$, V* the assumptions which we obtain from Assumptions I–V respectively by substituting $f^*(x, \xi)$ for $f(x, \theta)$, ξ for θ and $\hat{\xi}_n$ for $\hat{\theta}_n$. We shall show that Assumptions I*–V* can be derived from Assumptions I–VI.

Assumption I* is an immediate consequence of Assumptions I and VI. Since according to Assumption VI the first derivatives of $\xi^i(\theta)$ are continuous and bounded functions of θ, the transformation (135) is uniformly continuous. Hence, for each positive ϵ^* there exists a positive ϵ such that the inequality $|\hat{\theta}_n - \theta| \leqq \epsilon$ implies the inequality $|\hat{\xi}_n - \xi| \leqq \epsilon^*$. From this and Assumption II we obtain Assumption II*.

Denote by $\nu_i(x, \theta_1, \delta)$ the least upper bound, and by $\mu_i(x, \theta_1, \delta)$ the greatest lower bound of $\partial \log f(x, \theta)/\partial\theta^i$ in the interval $\theta_1 - \delta \leqq \theta \leqq \theta_1 + \delta$. Using the Taylor expansion we obtain

$$\frac{\partial \log f(x, \theta_1^*)}{\partial \theta^i} = \frac{\partial \log f(x, \theta_1)}{\partial \theta_1} + \sum_j (\theta_1^{*j} - \theta_1^j) \frac{\partial^2 \log f(x, \bar{\theta}_1)}{\partial\theta^i \partial\theta^j}, \tag{137}$$

where $\bar{\theta}_1$ lies in the interval $[\theta_1, \theta_1^*]$. From (137) it follows that

$$\left| \frac{\partial \log f(x, \theta_1^*)}{\partial \theta^i} - \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leqq 2\delta \sum_j [\,|\phi_{ij}(x, \theta, \delta)| + |\psi_{ij}(x, \theta, \delta)|\,] \tag{138}$$

for any θ and δ for which

$$\theta - \delta \leqq \theta_1 \leqq \theta + \delta \quad \text{and} \quad \theta - \delta \leqq \theta_1^* \leqq \theta + \delta.$$

From (138) we obtain for any positive δ

$$|\nu_i(x, \theta, \delta) - \mu_i(x, \theta, \delta)| \leqq 2\delta \sum_j [\,|\phi_{ij}(x, \theta, \delta)| + |\psi_{ij}(x, \theta, \delta)|\,]. \tag{139}$$

Let $\{\theta_{1n}\}$ and $\{\theta_{2n}\}$ $(n=1, 2, \cdots, \text{ad inf.})$ be two sequences of parameter points such that

$$\lim_{n=\infty} (\theta_{1n} - \theta_{2n}) = 0. \tag{140}$$

According to Assumption III the expectations $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded functions of θ_1, θ_2 and δ in the domain D_ϵ defined in Assumption III. Hence also the expectations $E_{\theta_1}|\phi_{ij}(x, \theta_2, \delta)|$ and $E_{\theta_1}|\psi_{ij}(x, \theta_2, \delta)|$ are bounded functions of θ_1, θ_2 and δ in the domain D_ϵ. From this and relations (139) and (140) it follows that for any sequence $\{\delta_n\}$ of positive numbers for which $\lim_{n=\infty} \delta_n = 0$ we have

$$\lim_{n=\infty} \{E_{\theta_{1n}}\nu_i(x, \theta_{2n}, \delta_n) - E_{\theta_{1n}}\mu_i(x, \theta_{2n}, \delta_n)\} = 0. \tag{141}$$

Since

$$\mu_i(x, \theta_{2n}, \delta_n) \leqq \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i} \leqq \nu_i(x, \theta_{2n}, \delta_n)$$

it follows from (141) that

$$\begin{aligned} &\lim_{n=\infty} \left\{E_{\theta_{1n}}\nu_i(x, \theta_{2n}, \delta_n) - E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i}\right\} = 0, \\ &\lim_{n=\infty} \left\{E_{\theta_{1n}}\mu_i(x, \theta_{2n}, \delta_n) - E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i}\right\} = 0. \end{aligned} \tag{142}$$

Using the Taylor expansion we have

$$\frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i} = \frac{\partial \log f(x, \theta_{1n})}{\partial \theta^i} + \sum_j (\theta^j_{2n} - \theta^j_{1n}) \frac{\partial^2 \log f(x, \bar{\theta}_n)}{\partial \theta^i \partial \theta^j}, \tag{143}$$

where $\bar{\theta}_n$ lies in the interval $[\theta_{1n}, \theta_{2n}]$. Since the expectations $E_{\theta_1}|\phi_{ij}(x, \theta_2, \delta)|$ and $E_{\theta_1}|\psi_{ij}(x, \theta_2, \delta)|$ are bounded functions of θ_1, θ_2, and δ in the domain D_ϵ, we obtain

$$\lim_{n=\infty} E_{\theta_{1n}}\left\{\sum_j (\theta^j_{2n} - \theta^j_{1n}) \frac{\partial^2 \log f(x, \bar{\theta}_n)}{\partial \theta^i \partial \theta^j}\right\} = 0.$$

Hence it follows from (143) and Assumption IV that

$$\lim_{n=\infty} E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i} = \lim_{n=\infty} E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{1n})}{\partial \theta^i} = 0. \tag{144}$$

We obtain from (142) and (144)

$$\lim E_{\theta_{1n}}\nu_i(x, \theta_{2n}, \delta_n) = \lim E_{\theta_{1n}}\mu_i(x, \theta_{2n}, \delta_n) = 0. \tag{145}$$

Denote by $\phi^*_{ij}(x, \xi_1, \delta^*)$ the least upper bound, and by $\psi^*_{ij}(x, \xi, \delta^*)$ the great-

est lower bound of $\partial^2 \log f^*(x, \xi)/\partial\xi^i\partial\xi^j$ in the interval $\xi_1-\delta^* \leqq \xi \leqq \xi_1+\delta^*$. We have

$$\frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i\partial\xi^j} = \sum_l \sum_m \frac{\partial^2 \log f(x, \theta)}{\partial\theta^l\partial\theta^m} \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} + \sum_l \frac{\partial \log f(x, \theta)}{\partial\theta^l} \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j}. \tag{146}$$

Since

$$E_\theta\left(\frac{\partial \log f(x, \theta)}{\partial\theta}\right) = 0,$$

we obtain from (146)

$$E_\xi \frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i\partial\xi^j} = \sum_l \sum_m E_\theta\left[\frac{\partial^2 \log f(x, \theta)}{\partial\theta^l\partial\theta^m} \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j}\right]. \tag{147}$$

Hence the determinant

$$\left| E_\xi \frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i\partial\xi^j} \right| = \left| E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i\partial\theta^j} \right| \left(\frac{\partial(\theta^1, \cdots, \theta^k)}{\partial(\xi^1, \cdots, \xi^k)}\right)^2. \tag{148}$$

Since the determinant $\left| -E_\theta\partial^2 \log f(x, \theta)/\partial\theta^i\partial\theta^j \right|$ has a positive lower bound. it follows from (148) and Assumption VI that

$$\left| - E_\xi\partial^2 \log f^*(x, \xi)/\partial\xi^i\partial\xi^j \right| \tag{149}$$

has a positive lower bound.

For any positive δ^* let $\delta(\delta^*)$ be the smallest positive number such that for any two points ξ_1 and ξ_2 for which $|\xi_1-\xi_2| \leqq \delta^*$ we have $|\theta_1-\theta_2| \leqq \delta(\delta^*)$ where θ_1 and θ_2 are the image points of ξ_1 and ξ_2 by transformation (136). From (146) we obtain

$$\begin{aligned} \phi^*_{ij}(x, \xi, \delta^*) &\leqq \sum_l \sum_m \bar{\phi}_{lm}[x, \theta, \delta(\delta^*)] \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} + \sum_l \bar{\nu}_l[x, \theta, \delta(\delta^*)] \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j}, \\ \psi^*_{ij}(x, \xi, \delta^*) &\geqq \sum_l \sum_m \bar{\psi}_{lm}[x, \theta, \delta(\delta^*)] \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} + \sum_l \bar{\mu}_l[x, \theta, \delta(\delta^*)] \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j}, \end{aligned} \tag{150}$$

where θ is the image point of ξ by transformation (136), and the derivatives $\partial\theta^l/\partial\xi^i$, $\partial^2\theta^l/\partial\xi^i\partial\xi^j$ are taken at some points in the interval $[\xi-\delta^*, \xi+\delta^*]$, and the functions $\bar{\phi}_{lm}(x, \theta, \delta)$, $\bar{\psi}_{lm}(x, \theta, \delta)$, $\bar{\mu}_l(x, \theta, \delta)$ and $\bar{\nu}_l(x, \theta, \delta)$ satisfy the inequalities

$$\psi_{lm}(x, \theta, \delta) \leqq \bar{\phi}_{lm}(x, \theta, \delta) \leqq \phi_{lm}(x, \theta, \delta);$$

$$\psi_{lm}(x, \theta, \delta) \leqq \bar{\psi}_{lm}(x, \theta, \delta) \leqq \phi_{lm}(x, \theta, \delta);$$

$$\mu_l(x, \theta, \delta) \leqq \bar{\nu}_l(x, \theta, \delta) \leqq \nu_l(x, \theta, \delta);$$

$$\mu_l(x, \theta, \delta) \leqq \underline{\mu}_l(x, \theta, \delta) \leqq \nu_l(x, \theta, \delta).$$

Let $\{\xi_{1n}\}$ and $\{\xi_{2n}\}$ be two sequences of parameter points such that $\lim \xi_{1n} = \lim \xi_{2n} = \xi$. Let θ_{1n} be the image point of ξ_{1n}, θ_{2n} the image point of ξ_{2n}, and θ the image point of ξ (by transformation (136)). Let furthermore $\{\delta_n^*\}$ be a sequence of positive numbers such that $\lim \delta_n^* = 0$. Then we obviously have $\lim \delta(\delta_n^*) = 0$ and therefore using (145) and Assumption III we obtain

$$\begin{aligned}
&\lim_{n=\infty} E_{\theta_{1n}} \phi_{lm}[x, \theta_{2n}, \delta(\delta_n^*)] = E_\theta \partial^2 \log f(x, \theta)/\partial\theta^l \partial\theta^m, \\
&\lim_{n=\infty} E_{\theta_{1n}} \psi_{lm}[x, \theta_{2n}, \delta(\delta_n^*)] = E_\theta \partial^2 \log f(x, \theta)/\partial\theta^l \partial\theta^m, \\
&\lim_{n=\infty} E_{\theta_{1n}} \nu_l \ [x, \theta_{2n}, \delta(\delta_n^*)] = 0, \\
&\lim_{n=\infty} E_{\theta_{1n}} \mu_l \ [x, \theta_{2n}, \delta(\delta_n^*)] = 0
\end{aligned} \tag{151}$$

uniformly in θ. From (150) and (151) and the uniform continuity of the derivatives $\partial\theta^l/\partial\xi^i$ and $\partial^2\theta^l/\partial\xi^i\partial\xi^j$ we obtain

$$\lim_{n=\infty} E_{\xi_{1n}} \phi^*_{ij}(x, \xi_{2n}, \delta_n^*) = \lim_{n=\infty} E_{\xi_{1n}} \psi^*_{ij}(x, \xi_{2n}, \delta_n^*) = E_\xi \frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i \partial\xi^j} \tag{152}$$

uniformly in ξ.

Because of (150) we have both

$$\begin{aligned}
&|\phi^*_{ij}(x, \xi, \delta^*)| \quad \text{and} \quad |\psi^*_{ij}(x, \xi, \delta^*)| \\
&\leqq \sum_l \sum_m \Big\{ [\,|\psi_{lm}[x, \theta, \delta(\delta^*)]| \\
&\qquad + |\phi_{lm}[x, \theta, \delta(\delta^*)]|\,] \Big[\text{l.u.b.}_\xi \Big| \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} \Big| \Big] \Big\} \\
&\quad + \sum_l \Big\{ [\,|\nu_l[x, \theta, \delta(\delta^*)]| + |\mu_l[x, \theta, \delta(\delta^*)]|\,] \Big[\text{l.u.b.}_\xi \Big| \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j} \Big| \Big] \Big\}
\end{aligned} \tag{153}$$

where the least upper bound with respect to ξ is to be taken over the interval $[\xi - \delta^*, \xi + \delta^*]$.

We shall show that $E_{\theta_1}[\nu_i(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\mu_i(x, \theta_2, \delta)]^2$ are bounded func-

tions of θ_1, θ_2 and δ in the domain D_ϵ for sufficiently small ϵ. Our statement is proved if we show that $E_{\theta_1}[\partial \log f(x, \theta_1)/\partial\theta^i]^2$, $E_{\theta_1}[\nu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ and $E_{\theta_1}[\mu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ are bounded in D_ϵ. The first of these expressions is bounded because of Assumption V. From (138) it follows that

$$\left| \nu_i(x, \theta_2, \delta) - \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leqq 2\delta' \sum_j [\,|\,\phi_{ij}(x, \theta_2, \delta')\,| + |\,\psi_{ij}(x, \theta_2, \delta')\,|\,]$$

and

$$\left| \mu_i(x, \theta_2, \delta) - \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leqq 2\delta' \sum_j [\,|\,\phi_{ij}(x, \theta_2, \delta')\,| + |\,\psi_{ij}(x, \theta_2, \delta')\,|\,]$$

where $\delta' = \delta + \sum_j |\theta_1^j - \theta_2^j|$. From the above inequalities and the fact that $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded in D_ϵ it follows that for sufficiently small ϵ the expressions $E_{\theta_1}[\nu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ and $E_{\theta_1}[\mu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^j]^2$ are bounded in D_ϵ. Hence our statement is proved. ∎

Since the derivatives $\partial\theta^i/\partial\xi^i$ and $\partial^2\theta^i/\partial\xi^i\partial\xi^j$ are bounded functions of ξ, and since $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded functions of θ_1, θ_2 and δ in the domain D_ϵ, it follows from (153) that there exists a positive ϵ^* such that $E_{\xi_1}[\phi_U^*(x, \xi_2, \delta^*)]^2$ and $E_{\xi_1}[\psi_U^*(x, \xi_2, \delta^*)]^2$ are bounded functions of ξ_1, ξ_2 and δ^* in the domain defined by $|\xi_1 - \xi_2| \leqq \epsilon^*$ and $|\delta^*| \leqq \epsilon^*$. Assumption III* follows from the latter statement and the relations (152) and (149).

Assumption IV* is an immediate consequence of Assumption IV.

We have

$$\frac{\partial \log f^*(x, \xi)}{\partial \xi^i} = \sum_j \frac{\partial \log f(x, \theta)}{\partial \theta^j} \frac{\partial \theta^j}{\partial \xi^i}. \tag{154}$$

For any points x and θ denote the maximum of the k expressions $|\partial \log f(x, \theta)/\partial\theta^1|, \cdots, |\partial \log f(x, \theta)/\partial\theta^k|$ by $\rho(x, \theta)$. From Assumption V it follows easily that

$$E_\theta[\rho(x, \theta)]^{2+\eta} \tag{155}$$

is a bounded function of θ. Since the derivatives $\partial\theta^j/\partial\xi^i$ are bounded functions of ξ, Assumption V* follows from (154) and (155).

Denote by $c_{ij}^*(\theta)$ the function of θ we obtain from $-E_\xi\partial^2 \log f^*(x, \xi)/\partial\xi^i\partial\xi^j$ by substituting $\xi(\theta)$ for ξ. Then we obtain from (147)

$$c_{ij}^*(\theta) = \sum_l \sum_m c_{lm}(\theta) \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j}. \tag{156}$$

Denote by A the matrix $\|\partial\theta^i/\partial\xi^j\|$ $(i, j = 1, \cdots, k)$ and let $\overline{A}$ be the trans-

posed of A. Then equation (156) can be written as

$$\|c^*_{ij}(\theta)\| = \overline{A}\|c_{ij}(\theta)\|A. \tag{157}$$

Let $\|\sigma^*_{ij}(\theta)\|$ be the inverse of the matrix $\|c^*_{ij}(\theta)\|$. Futhermore let B be the matrix $\|\partial\xi^i/\partial\theta^j\|$ $(i, j=1, \cdots, k)$ and denote by $\overline{B}$ the transposed of B. Since $B=A^{-1}$, we obtain from (157)

$$\|\sigma^*_{ij}(\theta)\| = B\|\sigma_{ij}(\theta)\|\overline{B}, \tag{158}$$

where $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$. Equation (158) can be written as

$$\sigma^*_{ij}(\theta) = \sum_l \sum_m \frac{\partial\xi^i}{\partial\theta^l}\frac{\partial\xi^j}{\partial\theta^m}\sigma_{lm}(\theta). \tag{159}$$

Let

$$\|c^*_{pq}(\theta)\| = \|\sigma^*_{pq}(\theta)\|^{-1} \qquad (p, q = 1, \cdots, r). \tag{160}$$

Denote by W_n^* the critical region defined by the inequality

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\theta_n)\xi^q(\theta_n)c^*_{pq}(\theta_n) \geqq d_n, \tag{161}$$

where the constant d_n is chosen so that

$$\text{l.u.b.}_{\theta}\ P(W^*\,|\,\theta) = \alpha.$$

The point θ is restricted to points of the set ω defined by equations (134).

For each positive c and for each point θ of ω we define the surface $S_c(\theta)$ by the equations

$$\begin{aligned}\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\theta)\xi^q(\theta)c^*_{pq}(\theta) &= c,\\ \sum_{j=1}^{k}\gamma_{tj}(\theta)\xi^j(\theta) &= \sum_{j=1}^{k}\gamma_{tj}(\theta)\xi^j(\theta) \qquad (t = r+1, \cdots, k),\end{aligned} \tag{162}$$

where the coefficients $\gamma_{tj}(\theta)$ satisfy the following condition: There exists a matrix $\|\beta_{pq}(\theta)\|$ $(p, q=1, \cdots, r)$ such that if we form the matrix $A(\theta)$ given in (120) then

$$A(\theta)\|\sigma^*_{ij}(\theta)\|\overline{A}(\theta) = I, \tag{163}$$

where $\overline{A}(\theta)$ denotes the transposed of $A(\theta)$ and I denotes the unit matrix.

Consider the transformation of the parameter space given by

$$\begin{aligned}\theta'^p &= \beta_{p1}(\theta)\xi^1(\theta) + \cdots + \beta_{pr}(\theta)\xi^r(\theta) \qquad (p = 1, \cdots, r),\\ \theta'^t &= \gamma_{t1}(\theta)\xi^1(\theta) + \cdots + \gamma_{tk}(\theta)\xi^k(\theta) \qquad (t = r+1, \cdots, k),\end{aligned} \tag{164}$$

where θ denotes the point of ω for which θ lies on $S_c(\theta)$ for some value of c. The matrix $\|\beta_{pq}(\theta)\|$ is chosen so that (163) is fulfilled. The transformation (164) transforms $S_c(\theta)$ into the sphere $S_c'(\theta)$ given by

$$\sum_{p=1}^{r} (\theta'^p)^2 = c, \qquad \theta'^i = \sum_{j=1}^{k} \gamma_{ij}(\theta)\xi^j(\theta).$$

We define a weight function $\eta(\theta)$ as follows:

$$\eta(\theta) = \lim_{\rho=0} A[\omega'(\theta, \rho)]/A[\omega(\theta, \rho)], \tag{165}$$

where the symbols on the right-hand side of (165) are defined as in (109).

Since Assumptions I*–V* are fulfilled if Assumption VI holds, we obtain from Theorems IV, V and VI the following theorem.

THEOREM VII. *Let W_n^* be the region defined in* (161) *and let $S_c(\theta)$ be the surface defined in* (162). *Furthermore let $\eta(\theta)$ be the weight function defined in* (165). *If Assumption* VI *holds, then for testing the hypothesis $\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$ the sequence $\{W_n^*\}$*

(a) *has asymptotically best average power with respect to the surfaces $S_c(\theta)$ and the weight function $\eta(\theta)$;*

(b) *has asymptotically best constant power on the surfaces $S_c(\theta)$;*

(c) *is an asymptotically most stringent test.*

13. **Optimum properties of the likelihood ratio test.** For testing a composite hypothesis H_ω, Neyman and Pearson introduced a statistic(5), called likelihood ratio, defined as follows: The density function in the sample space is given by $\prod_{\alpha=1}^{n} f(x_\alpha, \theta)$. Denote by $P(x_1, \cdots, x_n)$ the maximum of this function with respect to $\theta^1, \cdots, \theta^k$, and let $P_\omega(x_1, \cdots, x_n)$ be the conditional maximum with respect to $\theta^1, \cdots, \theta^k$, subject to the condition that θ must be a point of ω. Then the likelihood ratio for testing the hypothesis H_ω is given by $\lambda_n(\omega, E_n) = P_\omega(x_1, \cdots, x_n)/P(x_1, \cdots, x_n)$. It is obvious that the value of $\lambda_n(\omega, E_n)$ always lies between 0 and 1. Neyman and Pearson recommend the use of the left tail as critical region, that is the hypothesis H_ω is rejected if the value of $\lambda_n(\omega, E_n)$ is less than a certain constant $\bar{\lambda}_n(\omega)$. Denote the region $\lambda_n(\omega, E_n) < \bar{\lambda}_n(\omega)$ by $L_n(\omega)$. In all that follows we choose the constant $\bar{\lambda}_n(\omega)$ so that

$$\text{l.u.b.}_{\theta} P[L_n(\omega) \mid \theta] = \alpha.$$

We shall prove that there exists a finite value B such that $-2 \log \bar{\lambda}_n(\omega) < B$ for all n and for all ω. Consider the Taylor expansion

(5) See in this connection J. Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika vol. 20A (1928).

$$\sum_\alpha \log f(x_\alpha, \theta) = \sum_\alpha \log f(x_\alpha, \theta_n) + \frac{1}{2}\sum\sum n(\theta_n^i - \theta^i)(\theta_n^j - \theta^j) \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j} \frac{1}{n}$$

where $\bar{\theta}_n$ lies in the interval $[\theta_n, \theta]$. Since

$$\sum_\alpha \log f(x_\alpha, \theta_n) - \sum_\alpha \log f(x_\alpha, \theta) \geqq -\log \lambda_n(\omega, E_n),$$

we have

$$-2\log \lambda_n(\omega, E_n) \leqq -\sum\sum n(\theta_n^i - \theta^i)(\theta_n^j - \theta^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j}.$$

Since $n^{-1}\sum_\alpha \partial^2 \log f(x_\alpha, \bar{\theta}_n)/\partial\theta^i\partial\theta^j$ converges stochastically to $-c_{ij}(\theta)$ under the assumption that θ is the true parameter point, it follows easily from Proposition I that for any $\epsilon > 0$ there exists a positive value $A(\epsilon)$ such that for any ω

$$\limsup_{n=\infty} \left\{ \text{l.u.b.}_\theta P\left[-\sum\sum (\theta_n^i - \theta^i)(\theta_n^j - \theta^j) \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j} \geqq A(\epsilon) \,\middle|\, \theta \right] \right\} \leqq \epsilon$$

and $\lim_{\epsilon=0} A(\epsilon) = +\infty$. Hence

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta,\omega} P[-2\log \lambda_n(\omega, E_n) \geqq A(\epsilon) \mid \theta]\} \leqq \epsilon. \tag{166}$$

This proves the existence of a finite number B with the required property.

For any subsets Γ and Γ' of the parameter space we denote by $\delta(\Gamma, \Gamma')$ the greatest lower bound of the distance between θ and θ' where θ is restricted to points of Γ and θ' is restricted to points of Γ'. We shall call $\delta(\Gamma, \Gamma')$ the distance of the sets Γ and Γ'.

Let $\{\theta_n\}$ be a sequence of parameter points such that $\lim \delta(\theta_n, \omega) = 0$ and $\lim_{n=\infty} n^{1/2}\delta(\theta_n, \omega) = +\infty$. We shall prove that there exists a positive ν_0 such that for any constant A

$$\lim_{n=\infty} P\left[\sum_\alpha \log f(x_\alpha, \theta_n) - \sum_\alpha \log f(x_\alpha, \theta^*) > A \,\middle|\, \theta_n \right] = 1, \tag{167}$$

where θ^* denotes a point of ω for which $|\theta^* - \theta_n| \leqq \nu_0$ and $\sum_\alpha \log f(x_\alpha, \theta^*) \geqq \sum_\alpha \log f(x_\alpha, \theta)$ for all θ in the domain $|\theta - \theta_n| \leqq \nu_0$. Consider the Taylor expansion

$$\sum_\alpha \log f(x_\alpha, \theta^*) - \sum_\alpha \log f(x_\alpha, \theta_n) = \frac{n}{2}\sum\sum (\theta_n^i - \theta^{*i})(\theta_n^j - \theta^{*j}) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j}, \tag{168}$$

where $\bar{\theta}_n$ lies in the interval $[\hat{\theta}_n, \theta^*]$. Because of conditions (a) and (b) of Assumption III for any $\epsilon > 0$ the positive number ν_0 can be chosen so that

$$\lim_{n=\infty} P\left[\left|c_{ij}(\theta^*) + \frac{1}{n}\sum_{\alpha}\frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i\partial\theta^j}\right| < \epsilon \,\middle|\, \theta_n\right] = 1. \tag{169}$$

Since $\|c_{ij}(\theta)\|$ is positive definite and the determinant $|c_{ij}(\theta)|$ has a positive lower bound, we obtain from Proposition I, (168) and (169) that for some positive, ν_0, (167) holds. Hence our statement is proved.

We say that the likelihood ratio test is uniformly consistent if for any positive ν

$$\lim_{n=\infty} P[L_n(\omega) \mid \theta] = 1$$

uniformly in ω and θ over the domain $\delta(\theta, \omega) \geqq \nu$. We postulate the following assumption.

ASSUMPTION VII. *The likelihood ratio test is uniformly consistent.*

This assumption together with the uniform consistency of the maximum likelihood estimate $\hat{\theta}_n$ will be proved in a forthcoming paper on the basis of some weak assumptions on the density function $f(x, \theta)$.

Let $\omega_\nu(\theta)$ be the intersection of ω with the set of all points θ' for which $\sum_{i=1}^{k}|\theta^i - \theta'^i| \geqq \nu$. From Assumption VII it follows that for any positive ν

$$\lim_{n=\infty} P\{-2\log\lambda_n[\omega_\nu(\theta), E_n] \geqq -2\log\bar{\lambda}_n[\omega_\nu(\theta)] \mid \theta\} = 1 \tag{170}$$

uniformly in θ.

Let $\{\theta_n\}$ be a sequence of parameter points such that $\lim_{n=\infty}\delta(\theta_n, \omega) = 0$ and $\lim_{n=\infty} n^{1/2}\delta(\theta_n, \omega) = +\infty$. Denote by ω_n the set of all points θ for which $|\theta - \theta_n| \leqq \nu_0$. Since $-2\log\bar{\lambda}_n[\omega_\nu(\theta)]$ has a finite upper bound it follows from (167) that for a sufficiently small ν_0

$$\lim_{n=\infty} P\{-2\log\lambda_n[\omega_n, E_n] \geqq -2\log\bar{\lambda}_n[\omega_{\nu_0}(\theta_n)] \mid \theta_n\} = 1. \tag{171}$$

Obviously

$$-2\log\lambda_n(\omega, E_n) = \text{minimum}\{-2\log\lambda_n[\omega_{\nu_0}(\theta_n), E_n], -2\log\lambda_n(\omega_n, E_n)\}.$$

From (170) and (171) we obtain

$$\lim_{n=\infty} P\{-2\log\lambda_n(\omega, E_n) \geqq -2\log\bar{\lambda}_n[\omega_{\nu_0}(\theta_n)] \mid \theta_n\} = 1. \tag{172}$$

Since $-2\log\lambda_n[\omega_{\nu_0}(\theta_n), E_n] \geqq -2\log\lambda_n(\omega, E_n)$, we have $-2\log\bar{\lambda}_n[\omega_{\nu_0}(\theta_n)] \geqq -2\log\bar{\lambda}_n(\omega)$. Hence from (172) we obtain

$$\lim_{n=\infty} P[L_n(\omega) \mid \theta_n] = 1. \tag{173}$$

From Assumption VII it follows that (173) holds for any sequence $\{\theta_n\}$ for which $\lim n^{1/2}\delta(\theta_n, \omega) = +\infty$.

Now let us consider the case where ω is the set given by ${}_1\theta = {}_1\theta_0$. Let $\{\theta_n\}$ be a sequence of parameter points for which the sequence $n^{1/2}\delta(\theta_n, \omega)$ is bounded. Denote by $\hat{\theta}_n(\omega)$ the parameter point in ω for which

$$\prod_{\alpha=1}^{n} f[x_\alpha, \hat{\theta}_n(\omega)] = P_\omega(x_1, \cdots, x_n).$$

Let θ_n be the point for which $\delta(\theta_n, \omega) = \delta(\theta_n, \theta_n)$. Denote by $T_n(\nu)$ the region given by the inequalities

$$\begin{aligned} |\hat{\theta}_n - \theta_n| &< \nu, \\ |\hat{\theta}_n(\omega) - \theta_n| &< \nu. \end{aligned} \tag{174}$$

We shall prove that

$$\lim P[T_n(\nu) \mid \theta_n] = 1 \qquad \text{for any } \nu > 0. \tag{175}$$

Consider the Taylor expansion

$$\begin{aligned} \sum_\alpha \log f(x_\alpha, \theta_n) = \sum_\alpha \log f(x_\alpha, \hat{\theta}_n) \\ + \frac{n}{2} \sum\sum (\hat{\theta}_n^i - \theta_n^i)(\hat{\theta}_n^j - \theta_n^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j}. \end{aligned} \tag{176}$$

Since $[n^{-1}\sum_\alpha \partial^2 \log f(x_\alpha, \bar{\theta}_n)/\partial\theta^i\partial\theta^j + c_{ij}(\theta_n)]$ converges stochastically to zero, and since $n^{1/2}|\hat{\theta}_n^i - \theta_n^i|$ is bounded, we easily obtain from Proposition I that for any $\epsilon > 0$ there exists a constant B_ϵ such that

$$\limsup_{n=\infty} P[\sum \log f(x_\alpha, \hat{\theta}_n) - \sum \log f(x_\alpha, \theta_n) \geqq B_\epsilon \mid \theta_n] = \epsilon.$$

Since $-2 \log \lambda_n(\omega, E_n) \leqq 2\sum \log f(x_\alpha, \hat{\theta}_n) - 2\sum \log f(x_\alpha, \theta_n)$ we have

$$\limsup_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) \geqq 2B_\epsilon \mid \theta_n] \leqq \epsilon. \tag{177}$$

Denote by ω_n the subset of ω in which $|\theta - \theta_n| \geqq \nu$. From Assumption VII it follows that

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega_n, E_n) \geqq -2 \log \bar{\lambda}_n(\omega_n) \mid \theta_n] = 1. \tag{178}$$

Since $-2 \log \bar{\lambda}_n(\omega) \leqq -2 \log \bar{\lambda}_n(\omega_n)$ we have

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega_n, E_n) \geqq -2 \log \bar{\lambda}_n(\omega) \mid \theta_n] = 1. \tag{179}$$

For any given constant B there exists a positive $\alpha < 1$ such that $-2 \log \bar{\lambda}_n(\omega)$

$\geqq B$ if l.u.b.$_\theta P[L_n(\omega)|\theta]=\alpha$. Hence from (179) we obtain

$$\lim_{n=\infty} P[-2\log\lambda_n(\omega_n, E_n) \geqq B \,|\, \theta_n] = 1 \tag{180}$$

for any constant B. From (180) it follows that

$$\lim_{n=\infty} \{P[-2\log\lambda_n(\omega, E_n) \geqq B \,|\, \theta_n] - P[|\hat{\theta}_n(\omega) - \theta_n| \geqq \nu \,|\, \theta_n]\} \geqq 0 \tag{181}$$

for arbitrary values of B. From (181) and (177) we obtain

$$\lim_{n=\infty} P[|\hat{\theta}_n(\omega) - \theta_n| \geqq \nu \,|\, \theta_n] = 0.$$

Hence (175) is proved.

Consider the Taylor expansion

$$\begin{aligned}\sum_\alpha \log f(x_\alpha, \theta_n) &= \sum_\alpha \log f[x_\alpha, \hat{\theta}_n(\omega)] \\ &+ \frac{n}{2}\sum_{r+1}^{k}\sum_{r+1}^{k}[\theta_n^i(\omega) - \theta_n^i][\theta_n^j(\omega) - \theta_n^j]\,\frac{1}{n}\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i\partial\theta^j}\end{aligned} \tag{182}$$

where $\bar{\theta}_n$ lies in the interval $[\theta_n, \hat{\theta}_n(\omega)]$. In the following arguments we shall use the following lemma: *Let* $\|\lambda_{ij}\|$ $(i, j=1, \cdots, k)$ *be a definite matrix and for each integer* s *let* λ_{ij}^s *be a real number such that* $\lim_{s=\infty}\lambda_{ij}^s=\lambda_{ij}$. *Then*

$$\lim_{s=\infty}\left(\sum\sum \nu_i\nu_j\lambda_{ij} \Big/ \sum\sum \nu_i\nu_j\lambda_{ij}^s\right) = 1$$

uniformly in $\nu_1, \cdots, \nu_k$. From (175), the Taylor expansions (176) and (182), and the above lemma it follows that for any $\epsilon>0$

$$\begin{aligned}\lim_{n=\infty} P[&(1+\epsilon)q_n - (1-\epsilon)\bar{q}_n \geqq -2\log\lambda_n(\omega, E_n) \\ &\geqq (1-\epsilon)q_n - (1+\epsilon)\bar{q}_n \,|\, \theta_n] = 1,\end{aligned} \tag{183}$$

where

$$q_n = \sum_1^k\sum_1^k n(\theta_n^i - \theta_n^i)(\theta_n^j - \theta_n^j)c_{ij}(\theta_n) \tag{184}$$

and

$$\bar{q}_n = \sum_{r+1}^{k}\sum_{r+1}^{k} n[\theta_n^i(\omega) - \theta_n^i][\theta_n^j(\omega) - \theta_n^j]c_{ij}(\theta_n). \tag{185}$$

Since $-2\log\lambda_n(\omega, E_n)\geqq 0$, we obtain from (183)

$$\lim_{n=\infty} P[(1+\epsilon)q_n - (1-\epsilon)\bar{q}_n \geqq 0 \,|\, \theta_n] = 1. \tag{186}$$

Since the sequence $\{n^{1/2}\delta(\theta_n, \theta_n)\}$ is bounded, the expression q_n is bounded in the probability sense, that is for each positive ρ there exists a positive value A_ρ such that

$$\lim_{n=\infty} P(q_n > A_\rho \mid \theta_n) \leqq \rho.$$

From (186) it follows that $\bar{q}_n$ is also bounded in the probability sense. Hence because of (183) we have

$$\lim_{n=\infty} P[q_n - \bar{q}_n + \epsilon \geqq -2 \log \lambda_n(\omega, E_n) \geqq q_n - \bar{q}_n - \epsilon \mid \theta_n] = 1 \tag{187}$$

for any $\epsilon > 0$. From the Taylor expansions

$$\begin{aligned} y_n^i(\theta_n) &= \frac{1}{n^{1/2}} \sum_\alpha \frac{\partial \log f(x_\alpha, \theta_n)}{\partial \theta^i} \\ &= -\sum_{j=1}^{k} n^{1/2}(\theta_n^j - \theta_n^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta_n^*)}{\partial \theta^i \partial \theta^j} \qquad (i = 1, \cdots, k) \end{aligned} \tag{188}$$

and

$$y_n^i(\theta_n) = -\sum_{j=r+1}^{k} n^{1/2}[\theta_n^j(\omega) - \theta_n^j] \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta_n^{**})}{\partial \theta^i \partial \theta^j} \tag{189}$$

$$(i = r+1, \cdots, k)$$

we obtain

$$y_n^i(\theta_n) = +\sum_{j=1}^{k} n^{1/2}(\theta_n^j - \theta_n^j)[c_{ij}(\theta_n) + \epsilon_{ijn}(E_n)], \tag{190}$$

$$n^{1/2}(\theta_n^i - \theta_n^i) = \sum_{j=1}^{k} [\sigma_{ij}(\theta_n) + \eta_{ijn}(E_n)] y_n^j(\theta_n) \qquad (i = 1, \cdots, k), \tag{191}$$

and

$$n^{1/2}[\theta_n^i(\omega) - \theta_n^i] = \sum_{j=r+1}^{k} [\bar{\sigma}_{ij}(\theta_n) + \zeta_{ijn}(E_n)] y_n^j(\theta_n) \quad (i = r+1, \cdots, k), \tag{192}$$

where

$$\begin{aligned} \|\sigma_{ij}(\theta_n)\| &= \|c_{ij}(\theta_n)\|^{-1} \qquad (i, j = 1, \cdots, k), \\ \|\bar{\sigma}_{ij}(\theta)\| &= \|c_{ij}(\theta)\|^{-1} \qquad (i, j = r+1, \cdots, k), \end{aligned}$$

and for any positive ν we have

$$\begin{aligned} \lim_{n=\infty} P[|\epsilon_{ijn}(E_n)| < \nu \mid \theta_n] &= \lim_{n=\infty} P[|\eta_{ijn}(E_n)| < \nu \mid \theta_n] \\ &= \lim_{n=\infty} P[|\zeta_{ijn}(E_n)| < \nu \mid \theta_n] = 1. \end{aligned}$$

Hence

$$\sum_{r+1}^{k}\sum_{r+1}^{k} n[\theta_n^i(\omega) - \theta_n^i][\theta_n^j(\omega) - \theta_n^j]c_{ij}(\theta_n)$$

$$= \sum_{r+1}^{k}\sum_{r+1}^{k}\sum_{r+1}^{k}\sum_{r+1}^{k} [c_{ij}(\theta_n)\bar{\sigma}_{il}(\theta_n)\bar{\sigma}_{jm}(\theta_n) + \rho_{ijlmn}(E_n)]y_n^l(\theta_n)y_n^m(\theta_n)$$

$$= \sum_{r+1}^{k}\sum_{r+1}^{k} [\bar{\sigma}_{ij}(\theta_n) + \rho'_{ijn}(E_n)]y_n^i(\theta_n)y_n^j(\theta_n)$$

$$= \sum_{j=r+1}^{k}\sum_{i=r+1}^{k}\sum_{m=1}^{k}\sum_{l=1}^{k} [\bar{\sigma}_{ij}(\theta_n)c_{il}(\theta_n)c_{jm}(\theta_n) + \rho''_{ijmln}(E_n)]n(\theta_n^l - \theta_n^l)(\theta_n^m - \theta_n^m),$$

where

$$\lim_{n=\infty} P[|\rho''_{ijmln}(E_n)| < \nu | \theta_n] = 1$$

for any positive ν. If at least one of the integers l and m is greater than r, we have

$$\sum_{j=r+1}^{k}\sum_{i=r+1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{il}(\theta_n)c_{jm}(\theta_n) = c_{ml}(\theta_n).$$

Hence

$$\sum_{m=1}^{k}\sum_{l=1}^{k} v_m v_l c_{lm}(\theta_n) - \sum_{j=r+1}^{k}\sum_{i=r+1}^{k}\sum_{l=1}^{k}\sum_{m=1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{im}(\theta_n)c_{jm}(\theta_n)v_l v_m$$

$$= \sum_{m=1}^{r}\sum_{l=1}^{r} v_l v_m \left[c_{lm}(\theta_n) - \sum_{j=r+1}^{k}\sum_{i=r+1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{im}(\theta_n)c_{jl}(\theta_n) \right].$$

The coefficient

$$A_{lm}(\theta_n) = c_{lm}(\theta_n) - \sum_{r+1}^{k}\sum_{r+1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{im}(\theta_n)c_{jl}(\theta_n)$$

can be written as the following ratio:

$$A_{lm}(\theta_n) = \frac{\begin{vmatrix} c_{lm}(\theta_n) & c_{l\,r+1}(\theta_n) & \cdots & c_{lk}(\theta_n) \\ c_{m\,r+1}(\theta_n) & c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdot & \cdots & \cdot \\ c_{mk}(\theta_n) & c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix}}{\begin{vmatrix} c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdots & \cdot \\ c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix}}. \tag{193}$$

It is known that if A' is the adjoint of any determinant A, and M and M'

are corresponding m-rowed minors of A and A' respectively, then M' is equal to the product of A^{m-1} by the algebraic complement of M.

Let A be the k-rowed determinant $|\sigma_{ij}(\theta_n)|$ $(i, j=1, \cdots, k)$ and let M be the $(k-r+1)$-rowed minors

$$M = \begin{vmatrix} \sigma_{lm}(\theta_n) & \sigma_{l\,r+1}(\theta_n) & \cdots & \sigma_{lk}(\theta_n) \\ \sigma_{r+1\,m}(\theta_n) & \sigma_{r+1\,r+1}(\theta_n) & \cdots & \sigma_{r+1\,k}(\theta_n) \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{km}(\theta_n) & \sigma_{k\,r+1}(\theta_n) & \cdots & \sigma_{kk}(\theta_n) \end{vmatrix}. \tag{194}$$

Then we have

$$M' = A^{k-r+1} \begin{vmatrix} c_{lm}(\theta_n) & c_{l\,r+1}(\theta_n) & \cdots & c_{lk}(\theta_n) \\ c_{m\,r+1}(\theta_n) & c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdot & \cdots & \cdot \\ c_{mk}(\theta_n) & c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix} = A^{k-r}\overline{M}, \tag{195}$$

where $\overline{M}$ denotes the algebraic complement of M. Let M_1 be the $(k-r)$-rowed minor $|\sigma_{ij}(\theta)|$ $(i, j=r+1, \cdots, k)$. Then we obtain

$$M_1' = A^{k-r} \begin{vmatrix} c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdots & \cdot \\ c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix} = A^{k-r-1}\overline{M}_1, \tag{196}$$

where $\overline{M}_1$ denotes the algebraic complement of M_1.

From (193), (195) and (196) we obtain

$$A_{lm}(\theta_n) = \overline{M}/\overline{M}_1 = \bar{c}_{lm}(\theta_n) \qquad (l, m = 1, \cdots, r),$$

where $\|\bar{c}_{lm}(\theta_n)\| = \|\sigma_{lm}(\theta_n)\|^{-1}$ $(l, m=1, \cdots, r)$. Hence

$$q_n - \bar{q}_n = n\sum_{q=1}^{r}\sum_{p=1}^{r}(\theta_n^p - \theta_n^p)(\theta_n^q - \theta_n^q)[\bar{c}_{pq}(\theta_n) + \eta_{pqn}(E_n)] + \rho_n(E_n),$$

where for any positive ν

$$\lim_{n=\infty} P[|\eta_{pqn}(E_n)| < \nu \,|\, \theta_n] = \lim_{n=\infty} P[|\rho_n(E_n)| < \nu \,|\, \theta_n] = 1.$$

From (187) it follows that for any positive ϵ

$$\lim_{n=\infty} P\Big\{-2\log\lambda_n(\omega, E_n) - \epsilon \leqq \sum_{1}^{r}\sum_{1}^{r} n(\theta_n^p - \theta_n^p)(\theta_n^q - \theta_n^q) \cdot [\bar{c}_{pq}(\theta_n) + \eta_{pqn}(E_n)] \leqq -2\log\lambda_n(\omega, E_n) + \epsilon \,\Big|\, \theta_n\Big\} = 1. \tag{197}$$

Since $\theta_n^p = \theta_0^p$ $(p = 1, \cdots, r)$ from (197) we obtain because of the boundedness of the sequence $\{n^{1/2}(\theta_n - \theta_n)\}$

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leqq \sum\sum n(\theta_n^p - \theta_0^p)(\theta_n^q - \theta_0^q)\bar{c}_{pq}(\theta_n) \leqq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1.$$

The above equation remains valid if we substitute $\dot{\theta}_n$ for θ_n, that is

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leqq n\sum\sum(\dot{\theta}_n^p - \theta_0^p)(\dot{\theta}_n^q - \theta_0^q)\bar{c}_{pq}(\dot{\theta}_n) \leqq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1. \tag{198}$$

Let W_n^* be the critical region defined in (115). Since (173) holds for any sequence $\{\theta_n\}$ for which $\lim n^{1/2}\delta(\theta_n, \omega) = \infty$, we obtain from (198)

$$\lim_{n=\infty} \{P(W_n^* \mid \theta) - P[L_n(\omega) \mid \theta]\} = 0 \tag{199}$$

uniformly in θ.

Now we consider the general case where ω is given by r equations

$$\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$$

such that Assumption VI is satisfied. As we have seen in §12, the whole theory remains valid if we replace the parameters $\theta^1, \cdots, \theta^k$ by the new parameters

$$\xi^1 = \xi^1(\theta), \cdots, \xi^k = \xi^k(\theta),$$

where the functions $\xi^1(\theta), \cdots, \xi^k(\theta)$ satisfy Assumption VI. Hence from (199) it follows that

$$\lim_{n=\infty} \{P(W_n^* \mid \theta) - P[L_n(\omega) \mid \theta]\} = 0 \tag{200}$$

uniformly in θ, where W_n^* denotes the region defined in (161).

From (200) and Theorem VII we obtain the following theorem.

THEOREM VIII. *Let $S_c(\theta)$ be the surface defined in* (162) *and $\eta(\theta)$ be the weight function defined in* (165). *If Assumption* VI *holds, then for testing the hypothesis $\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$ the likelihood ratio test*

(a) *has asymptotically best average power with respect to the surfaces $S_c(\theta)$ and weight function $\eta(\theta)$;*

(b) *has asymptotically best constant power on the surfaces $S_c(\theta)$;*

(c) *is an asymptotically most stringent test.*

14. Large sample distribution of the likelihood ratio. S. S. Wilks(6) has derived the large sample distribution of the likelihood ratio $\lambda_n(\omega, E_n)$ if ω is

(6) S. S. Wilks, *Distribution of the likelihood ratio in large samples*, Ann. Math. Statist. 1938.

a linear subspace of the parameter space and if the hypothesis to be tested is true. Here we derive the large sample distribution of $\lambda_n(\omega, E_n)$ for any set ω satisfying Assumption VI in both cases, when the hypothesis to be tested is true, and when it is not true.

Let $u_1, \cdots, u_r$ be r independently and normally distributed variates with unit variances. Denote the expected value of u_p by μ_p. The distribution of the statistic

$$U^2 = u_1^2 + \cdots + u_r^2$$

is known(7). The only parameter involved in this distribution is $\lambda^2 = \mu_1^2 + \cdots + \mu_r^2$. Let us denote the cumulative distribution of U^2 by $F_r(\lambda^2, t)$, that is,

$$P[(U^2 < t)] = F_r(\lambda^2, t) \qquad (\lambda^2 = \mu_1^2 + \cdots + \mu_r^2). \tag{201}$$

Obviously $F_r(0, t)$ is the χ^2-distribution with r degrees of freedom.

Let $v_1, \cdots, v_r$ be r variates which have a joint normal distribution. Denote by μ_p the mean value of v_p and by σ_{pq} the covariance between v_p and v_q. Consider the statistic

$$V^2 = \sum_{q=1}^{r} \sum_{p=1}^{r} \lambda_{pq} v_p v_q, \tag{202}$$

where $\|\lambda_{pq}\| = \|\sigma_{pq}\|^{-1}$. It is easy to verify that the distribution of V^2 is given by

$$P(V^2 < t) = F_r(\lambda^2, t), \tag{203}$$

where

$$\lambda^2 = \sum \sum \lambda_{pq} \mu_p \mu_q. \tag{204}$$

We will now derive the limit distribution of the expression on the left-hand side of (161), that is of the statistic

$$Q_n = n \sum_{q=1}^{r} \sum_{p=1}^{r} \xi^p(\theta_n) \xi^q(\theta_n) c^*_{pq}(\theta_n). \tag{205}$$

The joint distribution of the variates $n^{1/2}[\xi^1(\theta_n) - \xi^1(\theta)], \cdots, n^{1/2}[\xi^r(\theta_n) - \xi^r(\theta)]$ converges with $n \to \infty$ uniformly towards the cumulative normal distribution with zero means and covariance matrix $\|\sigma^*_{pq}(\theta)\| = \|c^*_{pq}(\theta)\|^{-1}$. Since θ_n is a uniformly consistent estimate of θ and since $c_{ij}(\theta)$ is a uniformly continuous function of θ, the statistic

$$\overline{Q}_n = n \sum \sum \xi^p(\theta_n) \xi^q(\theta_n) c^*_{pq}(\theta) \tag{206}$$

(7) See for instance P. C. Tang, *The power function of the analysis of variance tests*, Statistical Research Memoirs vol. 2 (1938).

has the same limit distribution as Q_n, that is,

$$\lim_{n=\infty} \{P[\overline{Q}_n < t \mid \theta] - P[Q_n < t \mid \theta]\} = 0 \tag{207}$$

uniformly in θ and t.

It is easy to see that

$$\lim_{n=\infty} \{P[\overline{Q}_n < t \mid \theta] - F_r[\lambda_n^2(\theta), t] = 0 \tag{208}$$

uniformly in θ and t, where

$$\lambda_n^2(\theta) = n \sum \sum \xi^p(\theta)\xi^q(\theta)\bar{c}_{pq}^*(\theta). \tag{209}$$

Hence, because of (207) we have

$$\lim_{n=\infty} \{P[Q_n < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0 \tag{210}$$

uniformly in θ and t.

Let $\{\theta_n\}$ $(n=1, 2, \cdots, \text{ad inf.})$ be a sequence of parameter points for which $n^{1/2}\delta(\theta_n, \omega)$ is bounded. Then we obtain from (198)

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leqq Q_n \leqq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1 \tag{211}$$

for any positive ϵ. From (210) and (211) it follows that

$$\lim_{n=\infty} \{P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta_n] - F_r[\lambda_n^2(\theta_n), t]\} = 0 \tag{212}$$

uniformly in t. Since (173) holds for any sequence $\{\theta_n\}$ for which $\lim n^{1/2}\delta(\theta_n, \omega) = +\infty$, we obtain from (212)

$$\lim_{n=\infty} \{P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0 \tag{213}$$

uniformly in θ and t. Hence we have proved the following theorem.

THEOREM IX. *Let $F_r(\lambda^2, t)$ be the distribution function defined in* (201) *and let $\lambda_n(\omega, E_n)$ be the likelihood ratio statistic for testing the hypothesis $\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$. Let furthermore $\lambda_n^2(\theta)$ be the expression defined in* (209). *Then, if Assumption* VI *holds, we have*

$$\lim_{n=\infty} \{P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0$$

uniformly in t and θ. If the hypothesis to be tested is true, that is if θ is a point of ω, $\lambda_n^2(\theta) = 0$ and therefore the limit distribution of $-2 \log \lambda_n(\omega, E_n)$ is the χ^2-distribution with r degrees of freedom.

15. **Summary.** Let $f(x^1, \cdots, x^m, \theta^1, \cdots, \theta^k)$ be the joint probability density function of the variates $x^1, \cdots, x^m$ involving k unknown parameters $\theta^1, \cdots, \theta^k$. Any set of values $\theta^1, \cdots, \theta^k$ can be represented by a point θ of the k-dimensional Cartesian space with the coordinates $\theta^1, \cdots, \theta^k$. Let ω be the subset of the parameter space defined by the equations

$$\xi^1(\theta) = \xi^2(\theta) = \cdots = \xi^r(\theta) = 0 \qquad (r \leqq k),$$

that is, ω is the set of all points θ for which the above equations are fulfilled. Denote by H_ω the hypothesis that the true parameter point θ is an element of ω. In this paper the question of an appropriate test of the hypothesis H_ω is discussed when the number of observations is large.

The following notations have been introduced. The point $\dot{\theta}_n$ denotes the point with the coordinates $\theta_n^1, \cdots, \theta_n^k$ where θ_n^i is the maximum likelihood estimate of θ^i based on n independent observations on $x^1, \cdots, x^m$. The expected value of $-\partial^2 \log f(x^1, \cdots, x^m, \theta^1, \cdots, \theta^k)/\partial\theta^i\partial\theta^j$ is denoted by $c_{ij}(\theta)$ and $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$. Furthermore $\sum_{l=1}^{k}\sum_{m=1}^{k}(\partial\xi^p/\partial\theta^l)(\partial\xi^q/\partial\theta^m)\sigma_{lm}(\theta)$ $(p, q = 1, \cdots, r)$ is denoted by $\sigma_{pq}^*(\theta)$ and

$$\|\bar{c}_{pq}^*(\theta)\| = \|\sigma_{pq}^*(\theta)\|^{-1} \qquad (p, q = 1, \cdots, r).$$

The region W_n^* denotes the critical region defined by the inequality

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\dot{\theta}_n)\xi^q(\dot{\theta}_n)\bar{c}_{pq}^*(\dot{\theta}_n) \geqq d_n,$$

where n is the number of independent observations on $x^1, \cdots, x^m$, and the constant d_n is chosen so that the least upper bound of the probability that the sample point falls within W_n^*, calculated under the restriction that the true parameter point lies in ω, is equal to a given positive $\alpha < 1$.

Let λ_n be the likelihood ratio statistic for testing H_ω and let L_n be the critical region defined by the inequality

$$\lambda_n \leqq \bar{\lambda}_n,$$

where the constant $\bar{\lambda}_n$ is chosen so that the least upper bound of the probability that the sample point falls within L_n, calculated under the restriction that the true parameter point lies in ω, is equal to α.

Under certain assumptions on $f(x^1, \cdots, x^m, \theta^1, \cdots, \theta^k)$ and the functions $\xi^1(\theta), \cdots, \xi^r(\theta)$ the following results have been obtained:

I. For testing the hypothesis H_ω the critical regions W_n^* and L_n both: (1) have asymptotically best average power over a family of surfaces defined in (162); (2) have asymptotically best constant power along the surfaces defined in (162); (3) are asymptotically most stringent tests. The exact definitions of these notions are given in Definitions VIII, X and XII, respectively.

II. The statistics $-2 \log \lambda_n$ and $n\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\dot{\theta}_n)\xi^q(\dot{\theta}_n)\bar{c}_{pq}^*(\dot{\theta}_n)$ have the

same limit distribution. The limit distribution of $-2 \log \lambda_n$ is the χ^2-distribution with r degrees of freedom if the hypothesis to be tested is true. If the true parameter point θ_n is not an element of ω, the distribution of $-2 \log \lambda_n$ approaches the distribution of a sum of non-central squares

$$U^2 = u_1^2 + \cdots + u_r^2,$$

where the variates $u_1, \cdots, u_r$ are independently and normally distributed with unit variances and

$$\sum_{p=1}^{r} (Eu_p)^2 = n \sum \sum \xi^p(\theta_n) \xi^q(\theta_n) \bar{c}^*_{pq}(\theta_n).$$

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

ON SOME TRIGONOMETRIC SUMMABILITY METHODS AND GIBBS' PHENOMENON

BY

OTTO SZÁSZ

1. **Introduction.** We consider Fourier sine series

$$f(\theta) \sim \sum_1^\infty b_\nu \sin \nu\theta, \qquad \sum_1^n b_\nu \sin \nu\theta = s_n(\theta). \tag{1.1}$$

Suppose that $f(+0)$ exists and is greater than 0, that is $f(\theta)$ has a simple jump $2f(+0)$ at $\theta=0$. We say that the series (1.1) presents Gibbs' phenomenon, if

$$\limsup s_n(\theta_n) > f(+0), \text{ as } \theta_n \downarrow 0. \tag{1.2}$$

More generally, if only for some $k \geqq 0$

$$\lim_{\theta \to 0} \frac{2(k+1)}{\theta^{k+1}} \int_0^\theta (\theta - t)^k f(t)\,dt = j > 0$$

exists, and $\limsup s_n(\theta_n) > j/2$, we say that the series (1.1) presents a generalized Gibbs' phenomenon at $\theta=0$; j is the generalized jump of $f(\theta)$ at $\theta=0$. Our aim is to find general conditions for $f(\theta)$ or its Fourier coefficients, which imply a Gibbs' phenomenon. It is known that the jump j is closely connected with the asymptotic behavior of the sequence $\{nb_n\}$; on the other hand

$$s_n(\theta_n) = \sum_1^n \nu b_\nu \frac{\sin \nu\theta_n}{\nu}$$

is a linear transform of $\{nb_n\}$, with the triangular matrix $a_{n\nu} = \nu^{-1} \sin \nu\theta_n$, $\nu = 1, 2, \cdots, n$. In a previous paper [3](1) we have discussed the relationship of the transform

$$T_n(\theta_n) = \sum_1^n \tau_\nu \frac{\sin \nu\theta_n}{\nu}$$

to the Cesàro means of the sequence $\{\tau_n\}$. As an application we have proved the presence of a Gibbs' phenomenon under general assumptions and we gave new formulae for the determination of the jump $2f(+0)$. More general results are given in the present paper; the knowledge of (3) is not assumed.

We consider trigonometric linear transforms, related to T_n. We discuss in particular the transforms

Presented to the Society, April 24, 1943; received by the editors April 24, 1943.

(1) Numbers in brackets refer to the literature cited at the end of this paper.

$$B_n = \frac{1}{2}\{T_n(\theta_n) + T_n(\phi_n)\}, \qquad H_n = \frac{1}{n}\sum_{\nu=1}^{n} T_\nu(\theta_\nu),$$

$$S_n = \frac{1}{\theta_n}\int_0^{\theta_n} T_n(t)\,dt.$$

In §5 we prove a theorem on $(C, 2)$ summability of the sequence $\{nb_n\}$ for Fourier series.

We shall make repeated use of the well known theorem:

THEOREM A. *The convergence of a sequence* $\{\tau_n\}$ *implies the convergence of the transform* $T_n = \sum_1^n a_{n\nu}\tau_\nu$, *if and only if*

$$\lim_{n\to\infty} a_{n\nu} = a_\nu \text{ exists for } \nu = 1, 2, 3, \cdots,$$

$$\sum_{\nu=1}^{n} |a_{n\nu}| = O(1), \quad \text{as} \quad n \to \infty,$$

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} a_{n\nu} = \sigma \text{ exists.}$$

We then have $\lim T_n = \sigma \lim \tau_n + \sum_1^\infty a_\nu(\tau_\nu - \lim \tau_n)$(*).

2. **The average** $(1/2)\{T_n(\theta_n)+T_n(\phi_n)\} = B_n$. On putting $\rho_n = 1$ in Theorem 2 of our paper [3] we get easily the following theorem.

THEOREM 1. *Suppose that*

$$(2.1) \qquad \frac{1}{n}\sum_{1}^{n} \tau_\nu \to \tau, \quad \phi_n > 0, \quad \theta_n \geqq 0, \quad n\theta_n = O(1), \quad n\phi_n = O(1),$$

and that $\sigma(\alpha) = \int_0^\alpha (t^{-1}\sin t)\,dt$ *has the same value for all limit points* α *of the sequence* $\{n\theta_n\}$, *and the same value for all limit points* β *of the sequence* $\{n\phi_n\}$; *then*

$$(2.2) \qquad \lim_{n\to\infty} B_n = (\tau/2)(\sigma(\alpha) + \sigma(\beta)).$$

For the special case $\phi_n = \pi/n$ see Rogosinski [1, 2].

We next assume $(C, 2)$ $\lim \tau_n = \tau$. Let $\tau_n' = \sum_1^n \tau_\nu$, $\tau_n^2 = \sum_1^n \tau_\nu^1$, then our assumption is: $2\tau_n^2/n^2 \to \tau$. We write $\Delta^0\tau_n = \tau_n$, $\Delta\tau_n = \tau_n - \tau_{n+1}$, $\Delta^2\tau_n = \Delta(\Delta\tau_n)$; then

$$(2.3) \qquad B_n = \sum_{1}^{n-2} {}^2\tau_\nu \Delta^2 \frac{\sin \nu\theta_\nu + \sin \nu\phi_\nu}{\nu} + {}^2\tau_{n-1}\left\{\frac{\sin (n-1)\theta_n + \sin (n-1)\phi_n}{n-1} - 2\frac{\sin n\theta_n + \sin n\phi_n}{n}\right\} + {}^2\tau_n \frac{\sin n\theta_n + \sin n\phi_n}{n}.$$

(*) The statement at the beginning of §2 in [3] should be corrected accordingly; it does not affect the rest of the paper.

It follows from Theorem A that, if in addition to the assumptions of Theorem 1,

$$\sum_1^{n-2} \nu^2 \left| \Delta^2 \frac{\sin \nu\theta_n + \sin \nu\phi_n}{\nu} \right| + n^2 \left| \Delta \frac{\sin (n-1)\theta_n + \sin (n-1)\phi_n}{n-1} \right| + n \left| \sin n\theta_n + \sin n\phi_n \right| = O(1),$$

then again (2.2) holds.

We now assume $n(\phi_n - \theta_n) = (2\mu - 1)\pi + \delta_n$, $\delta_n = O(1/n)$, μ an integer $\geqq 1$. Then

$$\sin n\theta_n + \sin n\phi_n = 2 \sin \frac{n}{2} (\theta_n + \phi_n) \cos \frac{n}{2} (\theta_n - \phi_n) = O\left(\frac{1}{n}\right);$$

also

$$\cos \frac{n-1}{2} (\theta_n - \phi_n) = \cos \frac{n}{2} (\theta_n - \phi_n) \cos \frac{\theta_n - \phi_n}{2} + \sin \frac{n}{2} (\theta_n - \phi_n) \sin \frac{\theta_n - \phi_n}{2} = O\left(\frac{1}{n}\right).$$

Furthermore, for $\theta > 0$,

$$\Delta^2 \left(\frac{\sin \nu\theta}{\nu} \right) = \Delta^2 \int_0^\theta \cos \nu t dt = R \int_0^\theta \Delta^2 z^\nu dt, \qquad z = e^{it}.$$

Hence

$$\Delta^2 \frac{\sin \nu\theta}{\nu} = R \int_0^\theta z^\nu (1 - z)^2 dt,$$

and

$$\left| \Delta^2 \frac{\sin \nu\theta}{\nu} \right| < \int_0^\theta |1 - z|^2 dt = \int_0^\theta \left| e^{-it/2} - e^{it/2} \right|^2 dt \tag{2.4}$$

$$= 4 \int_0^\theta \sin^2 \frac{t}{2} dt < \int_0^\theta t^2 dt = \frac{\theta^3}{3}.$$

Thus in view of (2.1)

$$\sum_1^{n-2} \nu^2 \left| \Delta^2 \frac{\sin \nu\theta_n + \sin \nu\phi_n}{\nu} \right| < (\theta_n^3 + \phi_n^3) \sum_1^n \nu^2 = O(1).$$

We have thus proved the following theorem.

THEOREM 2. *Suppose that* $(C, 2)$ $\lim \tau_n = \tau$, $\phi_n > 0$, $\theta_n \geqq 0$, $n\theta_n = O(1)$, $n\phi_n = n\theta_n + (2\mu - 1)\pi + O(1/n)$, μ *an integer greater than or equal to* 1, *and that* $\sigma(\alpha) = \int_0^\alpha t^{-1} \sin t dt$ *has the same value for all limit points* α *of the sequence* $\{n\theta_n\}$. *Then* $\lim B_n = (\tau/2)\{\sigma(\alpha) + \sigma(\alpha + (2\mu - 1)\pi)\}$.

3. **The average** $(1/n) \sum_{\nu=1}^n T_\nu(\theta_\nu) = H_n$. We have

$$(3.1)\qquad nH_n = \sum_{\nu=1}^{n}\sum_{\mu=1}^{\nu} \nu^{-1}\tau_\nu \sin \mu\theta_\mu = \sum_{\nu=1}^{n} \nu^{-1}\tau_\nu \sum_{\mu=\nu}^{n} \sin \mu\theta_\mu = \sum_{\nu=1}^{n} C_{n\nu}\tau_\nu,$$

where

$$(3.2)\qquad C_{n\nu} = \nu^{-1}\sum_{\mu=\nu}^{n} \sin \mu\theta_\mu, \qquad \nu = 1, 2, \cdots, n.$$

Thus $nH_n = \sum_1^{n-1}(C_{n\nu} - C_{n,\nu+1})\tau_\nu + C_{nn}\tau_n'$. We first prove the following theorem.

THEOREM 3. *If* $(C, 1)$ $\lim \tau_n = \tau$, *and*

$$(3.3)\qquad \frac{1}{n}\sum_{1}^{n} |\nu\theta_\nu - \alpha| \to 0,$$

then

$$(3.4)\qquad H_n \to \tau \int_0^\alpha \frac{\sin t}{t}\, dt.$$

(3.3) is strong summability $(C, 1)$ of the sequence $\{n\theta_n\}$ to α; it implies that $\theta_n \to 0$, hence $\lim_{n\to\infty} n^{-1}C_{n\nu} = 0$ for each ν. Furthermore

$$\begin{aligned}
&\sum_{1}^{n-1} \nu |C_{n\nu} - C_{n,\nu+1}| + n|C_{nn}| \\
&\quad = \sum_{1}^{n-1} \left| \sum_{\mu=\nu}^{n} \sin \mu\theta_\mu - \frac{\nu}{\nu+1}\sum_{\mu=\nu+1}^{n} \sin (\nu+1)\theta_\mu \right| + |\sin n\theta_n| \\
&\quad < \sum_{1}^{n-1} |\sin \nu\theta_\nu| + \sum_{1}^{n-1} \left| \sum_{\mu=\nu+1}^{n} (\sin \mu\theta_\mu - \sin (\nu+1)\theta_\mu) \right| \\
&\qquad + \sum_{1}^{-1} \frac{1}{\nu+1} \left| \sum_{\mu=\nu+1}^{n} \sin (\nu+1)\theta_\mu \right| + 1 \\
&\quad < \sum_{1}^{n} \nu\theta_\nu + 2\sum_{1}^{n-1}\sum_{\mu=\nu+1}^{n} \left| \sin \frac{\theta_\mu}{2} \cos \frac{2\nu+1}{2}\theta_\mu \right| + \sum_{1}^{n-1}\sum_{\mu=\nu+1}^{n} \theta_\mu + 1 \\
&\quad < 3\sum_{1}^{n} \nu\theta_\nu + 1 = O(n), \quad \text{as} \quad n \to \infty,
\end{aligned}$$

hence $H_n = O(1)$. Thus, in order that H_n has a limit whenever $n^{-1}\tau_n'$ tends to a limit τ, the additional condition is that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{\nu=1}^{n} C_{n\nu} = \lim \frac{1}{n}\sum_{\nu=1}^{n} \nu^{-1}\sum_{\mu=\nu}^{n} \sin \mu\theta_\mu = \sigma$$

exists. We then have $H_n \to \sigma\tau$ (cf. Theorem A). Now

$$\frac{1}{n}\sum_{1}^{n} C_{n\nu} = \frac{1}{n}\sum_{\kappa=1}^{n} S_\kappa(\theta_k),$$

and

$$S_\kappa(\theta_\kappa) = \sum_{\lambda=1}^{\kappa} \lambda^{-1} \sin \lambda\theta_\kappa = \int_0^{\theta_\kappa} \left(\sum_{1}^{\kappa} \cos \lambda t \right) dt$$

$$= -\frac{\theta_\kappa}{2} + \int_0^{\theta_\kappa} \frac{\sin (\kappa + 1/2)t}{2 \sin (t/2)}\, dt.$$

Furthermore

$$\int_0^{\theta} \frac{\sin (\kappa + 1/2)t}{2 \sin (t/2)}\, dt = \int_0^{\theta} \frac{\sin (\kappa + 1/2)t}{t}\, dt$$

$$+ \int_0^{\theta} \sin (\kappa + 1/2)t \left\{ \frac{1}{2 \sin (t/2)} - \frac{1}{t} \right\} dt,$$

and, from the mean value theorem,

$$(3.5) \qquad 0 < \frac{1}{2 \sin (t/2)} - \frac{1}{t} = \frac{t - 2 \sin (t/2)}{2t \sin (t/2)} < \frac{t}{12}, \qquad \theta < t < \pi.$$

Hence

$$\left| \int_0^{\theta} \sin (\kappa + 1/2)t \left\{ \frac{1}{2 \sin (t/2)} - \frac{1}{t} \right\} dt \right| < \frac{1}{12} \int_0^{\theta} t\, dt = \frac{1}{24}\theta^2,$$

and

$$S_\kappa(\theta_\kappa) = -\frac{\theta_\kappa}{2} + \int_0^{(\kappa+1/2)\theta_\kappa} \frac{\sin t}{t}\, dt + \delta_\kappa \theta_\kappa^2$$

$$= \int_0^{\alpha} \frac{\sin t}{t}\, dt + \int_{\alpha}^{(\kappa+1/2)\theta_\kappa} \frac{\sin t}{t}\, dt - \frac{\theta_\kappa}{2} + \delta_\kappa \theta_\kappa^2,$$

where $|\delta_\kappa| < 1/24$. Thus, for $0 < \theta_\kappa < 1$,

$$\left| S_\kappa(\theta_\kappa) - \int_0^{\alpha} \frac{\sin t}{t}\, dt \right| < |(\kappa + 1/2)\theta_\kappa - \alpha| + \theta_\kappa < |\kappa\theta_\kappa - \alpha| + 2\theta_\kappa,$$

and from (3.3)

$$\left| \frac{1}{n}\sum_{1}^{n} S_\kappa(\theta_\kappa) - \int_0^{\alpha} \frac{\sin t}{t}\, dt \right| < \frac{1}{n}\sum_{1}^{n} |\kappa\theta_\kappa - \alpha| + \frac{2}{n}\sum_{1}^{n} \theta_\kappa \to 0;$$

which proves Theorem 3. For the case of $(C, 2)$ summability we prove the following theorem.

THEOREM 4: *If* $(C, 2)$ $\lim \tau_n = \tau$,

$$\sum_1^n \nu^2 |\theta_\nu - \theta_{\nu+1}| = O(n), \quad \text{as} \quad n \to \infty, \tag{3.6}$$

and if (3.3) *holds, then* (3.4) *holds.*

Evidently (3.3) implies $\sum_1^n \nu\theta_\nu = O(n)$; furthermore

$$\sum_1^n \nu^2(\theta_\nu - \theta_{\nu+1}) = \sum_1^n (2\nu - 1)\theta_\nu - n^2\theta_{n+1} = O(n) - n^2\theta_{n+1}, \tag{3.7}$$

hence from (3.6)

$$n\theta_n = O(1) \quad \text{as} \quad n \to \infty. \tag{3.8}$$

We have from (3.1)

$$nH_n = \sum_1^{n-2} \tau_\nu'' \Delta^2 C_{n\nu} + \tau_{n-1}''(C_{n,n-1} - 2C_{nn}) + \tau_n'' C_{nn},$$

where $nC_{nn} = \sin n\theta_n = O(1)$, and

$$n(C_{n,n-1} - 2C_{nn}) = O(1) + (n/(n-1))\{\sin(n-1)\theta_{n-1} + \sin(n-1)\theta_n\} = O(1).$$

Furthermore, using (2.4) and (3.8),

$$\begin{aligned}
&\sum_1^{n-2} \nu^2 |\Delta^2 C_{n\nu}| \\
&= \sum_1^{n-2} \nu^2 \left| \frac{1}{\nu} \sum_{\kappa=\nu}^{n} \sin \nu\theta_\kappa - \frac{2}{\nu+1} \sum_{\kappa=\nu+1}^{n} \sin(\nu+1)\theta_\kappa + \frac{1}{\nu+2} \sum_{\kappa=\nu+2}^{n} \sin(\nu+2)\theta_\kappa \right| \\
&\leqq \sum_1^{n-2} \nu^2 \left| \frac{1}{\nu}(\sin \nu\theta_\nu + \sin \nu\theta_{\nu+1}) - \frac{2}{\nu+1} \sin(\nu+1)\theta_{\nu+1} \right| \\
&\qquad + \sum_1^{n-2} \nu^2 \left| \sum_{\kappa=\nu+2}^{n} \left(\frac{\sin \nu\theta_\kappa}{\nu} - \frac{2 \sin(\nu+1)\theta_\kappa}{\nu+1} + \frac{\sin(\nu+2)\theta_\kappa}{\nu+2} \right) \right| \\
&< O(n) + \sum_1^{n-1} \nu |\sin \nu\theta_\nu + \sin \nu\theta_{\nu+1} - 2 \sin(\nu+1)\theta_{\nu+1}| \\
&\qquad + \sum_1^{n-2} \nu^2 \sum_{\kappa=\nu+2}^{n} |\Delta^2 \nu^{-1} \sin \nu\theta_\kappa| \\
&< O(n) + \sum_1^{n-1} \nu(\theta_{\nu+1} + \theta_{\nu+1} + \nu|\theta_\nu - \theta_{\nu+1}|) + \sum_1^{n} \nu^2 \left(\sum_{\kappa=\nu}^{n} \theta_\kappa^2 \right) \\
&= O(n) + O\left(\sum_1^{n} \nu^3 \theta_\nu^3 \right) = O(n).
\end{aligned}$$

Now (3.4) follows as in Theorem 3. This proves Theorem 4. In this connection the following lemma is of interest:

LEMMA. *If for some* $c>0$

$$0 \leqq \theta_{n+1} \leqq (1 + c/n)\theta_n, \qquad n = 1, 2, 3, \cdots, \tag{3.9}$$

and

$$\sum_1^n \nu\theta_\nu = O(n), \qquad \text{as } n \to \infty, \tag{3.10}$$

then (3.6) *and* (3.8) *hold.*

We write $\sum_1^n \nu^2|\theta_\nu - \theta_{\nu+1}| = \sum' + \sum''$, where $\sum'$ contains all terms with $\theta_\nu \leqq \theta_{\nu+1}$, and $\sum''$ the rest. Now, using (3.9) and (3.10),

$$\sum' \leqq c\sum_1^n \nu\theta_\nu = O(n),$$

and from (3.7)

$$\sum'' - \sum' < \sum_1^n (2\nu - 1)\theta_\nu = O(n),$$

hence $\sum'' = O(n)$ and (3.6) is proved. This and (3.10) yield (3.8). In particular in Theorem 4 assumption (3.6) can be replaced by $\theta_n \downarrow$.

4. **The integral mean** $(1/\theta_n)\int_0^{\theta_n}(\sum_1^n \tau_\nu \nu^{-1} \sin \nu t)dt$. To a given sequence $\{\tau_n\}$ we consider the transform

$$S_n = \frac{1}{\theta_n}\int_0^{\theta_n}\left(\sum_1^n \tau_\nu \sin \nu t\right)dt = \sum_1^n \frac{1 - \cos \nu\theta_n}{\nu^2\theta_n}\tau_\nu = \sum_1^n a_{n\nu}\tau_\nu, \quad \theta_n \to 0; \tag{4.1}$$

so that now

$$a_{n\nu} = (1 - \cos \nu\theta_n)/\nu^2\theta_n, \qquad \nu = 1, 2, \cdots, n. \tag{4.2}$$

Evidently $0 < a_{n\nu} \to 0$ as $n \to \infty$; hence a necessary and sufficient condition that for every convergent sequence $\tau_n \to \tau$ the sequence S_n has a limit $\sigma\tau$ is that $\lim \sum_1^n A_{n\nu} = \sigma$ exists (from Theorem A). Now

$$\begin{aligned}\sum_1^n \frac{1 - \cos \nu\theta_n}{\nu^2} &= \int_0^{\theta_n}\left(\sum_1^n \frac{\sin \nu t}{\nu}\right)dt \\ &= \int_0^{\theta_n}\left\{-\frac{t}{2} + \int_0^t \frac{\sin (n + 1/2)u}{2 \sin (u/2)}\,du\right\}dt \\ &= -\frac{\theta_n^2}{4} + \int_0^{\theta_n}(\theta_n - t)\frac{\sin (n + 1/2)t}{2 \sin (t/2)}\,dt,\end{aligned}$$

thus

$$\sum_1^n a_{n\nu} = -\frac{\theta_n^2}{4} + \theta_n^{-1}\int_0^{\theta_n}(\theta_n - t)\frac{\sin(n+1/2)t}{t}\,dt$$
$$+ \theta_n^{-1}\int_0^{\theta_n}(\theta_n - t)\left\{\frac{1}{2\sin(t/2)} - \frac{1}{t}\right\}\sin(n+1/2)t\,dt,$$

and using (3.5)

$$\theta_n^{-1}\left|\int_0^{\theta_n}(\theta_n - t)\left\{\frac{1}{2\sin(t/2)} - \frac{1}{t}\right\}\sin(2n+1)\frac{t}{2}\,dt\right|$$
$$< \theta_n^{-1}\int_0^{\theta_n}(\theta_n - t)\frac{t\,dt}{12} < \frac{1}{24}\theta_n^2 \to 0.$$

Hence

$$\sum_1^n a_{n\nu} = o(1) + \theta_n^{-1}\int_0^{(n+1/2)\theta_n}\left(\theta_n - \frac{t}{n+1/2}\right)\frac{\sin t}{t}\,dt$$
$$= o(1) + \int_0^{(n+1/2)\theta_n}\frac{\sin t}{t}\,dt - \frac{2}{(2n+1)\theta_n}[1 - \cos(n+1/2)\theta_n].$$

If $\beta \leqq \infty$ is a limit point of the sequence $\{n\theta_n\}$, then for a subsequence of integers n

$$\sum_1^n a_{n\nu} \to \int_0^{\beta}\frac{\sin t}{t}\,dt - \frac{1-\cos\beta}{\beta} = g(\beta).$$

Now $g'(\beta) = (1-\cos\beta)/\beta^2 \geqq 0$, hence $g(\beta) \uparrow$, as β increases, to $g(\infty) = \pi/2$. We have thus proved:

THEOREM 5. *In order that for every convergent sequence $\tau_n \to \tau$ the transform $S_n = \sum_1^n \tau_\nu(1-\cos\nu\theta_n)/\nu^2\theta_n$ has a limit, it is necessary and sufficient that $\lim_{n\to\infty}\theta_n = \beta$, $\beta \leqq \infty$, exists. We then have $S_n \to \tau g(\beta)$, where $g(\beta) = \int_0^\beta (t^{-1}\sin t)dt - (1-\cos\beta)/\beta \leqq \pi/2$.*

We now assume only $n^{-1}\sum_1^n \tau_\nu \to \tau$, or $n^{-1}\tau_n' \to \tau$. We have $S_n = \sum_1^{n-1}(a_{n\nu} - a_{n,\nu+1})\tau_\nu' + a_{nn}\tau_n'$; and the additional condition for the existence of $\lim S_n$ is by Theorem A:

$$\sum_1^{n-1}\nu\left|\frac{1-\cos\nu\theta_n}{\nu^2\theta_n} - \frac{1-\cos(\nu+1)\theta_n}{(\nu+1)^2\theta_n}\right| + \frac{1-\cos n\theta_n}{n\theta_n} = O(1),$$

or

$$(4.3)\qquad \theta_n^{-1}\sum_1^{n-1}\nu^{-3}\left|(\nu+1)^2(1-\cos\nu\theta_n) - \nu^2(1-\cos(\nu+1)\theta_n)\right| \leqq O(1).$$

But

$$\theta_n^{-1}\sum_1^n \nu^{-3}(2\nu+1)(1-\cos\nu\theta_n) < 3\sum_1^n \frac{1-\cos\nu\theta_n}{\nu^2\theta_n} = 3\sum_1^n a_{n\nu} = O(1),$$

hence (4.3) reduces to $\theta_n^{-1}\sum_1^{n-1}\nu^{-1}|\cos\nu\theta_n-\cos(\nu+1)\theta_n| = O(1)$, as $n\to\infty$, or $\sum_1^n\nu^{-1}|\sin(\nu+1/2)\theta_n| = O(1)$.

But this condition is equivalent to $n\theta_n = O(1)$ (cf. [3, §2]). Thus we have the following:

THEOREM 6. *The transform* $S_n=\sum_1^n\tau_\nu(1-\cos\nu\theta_n)/\nu^2\theta_n$ *has a limit for every sequence* $\{\tau_n\}$ *which is summable* $(C, 1)$ *to* τ, *if and only if* $\lim n\theta_n=\beta<\infty$ *exists. We then have* $S_n\to\tau g(\beta)$.

We finally assume $(C, 2)$ $\lim\tau_n=\tau$; using the formula

$$\sum_1^n a_\nu\tau_\nu = \sum_1^{n-2}\tau_\nu^2\Delta^2 a_\nu + \tau_{n-1}^2(a_{n-1}-2a_n) + \tau_n^2 a_n,$$

we now get for the existence of $\lim S_n$ the additional conditions

$$(4.4)\qquad \theta_n^{-1}\sum_1^{n-2}\nu^2\left|\frac{1-\cos\nu\theta_n}{\nu^2} - 2\frac{1-\cos(\nu+1)\theta_n}{(\nu+1)^2} + \frac{1-\cos(\nu+2)\theta_n}{(\nu+2)^2}\right| = O(1),$$

$$(4.5)\qquad \theta_n^{-1}n^2\left|\frac{1-\cos(n-1)\theta_n}{(n-1)^2} - 2\frac{1-\cos n\theta_n}{n^2}\right| + \frac{1-\cos n\theta_n}{\theta_n} = O(1).$$

In particular $\sin^2 n\theta_n/2 = O(\theta_n) = o(1)$, hence $n\theta_n\to\beta=2\lambda\pi$, λ an integer. We assume $\lambda>0$; on putting $n\theta_n=2\lambda\pi+2\epsilon_n$, $\epsilon_n\to0$, we must have

$$\sin^2(\lambda\pi+\epsilon_n) = \sin^2\epsilon_n = O((\lambda\pi+\epsilon_n)/n) = O(1/n),$$

or $\epsilon_n=O(1/n^{1/2})$. To satisfy (4.5) the additional condition is $\sin^2(n-1)\theta_n = O(\theta_n)$, or $(\sin n\theta_n\cos\theta_n-\cos n\theta_n\sin\theta_n)^2=O(1/n)$, which reduces to our previous condition. We finally show that now (4.4) is also satisfied. We have

$$\frac{1-\cos\nu\theta}{\nu^2} = \int_0^\theta dt\int_0^t\cos\nu u\,du = \int_0^\theta(\theta-t)\cos\nu t\,dt,$$

hence

$$\Delta^2\frac{1-\cos\nu\theta}{\nu^2} = \int_0^\theta(\theta-t)\Delta^2\cos\nu t\,dt = R\int_0^\theta(\theta-t)z^\nu(1-z)^2dt,\quad z=e^{it},$$

and

$$\left|\Delta^2\frac{1-\cos\nu\theta}{\nu^2}\right| < \int_0^\theta(\theta-t)\,|1-e^{it}|^2dt < \int_0^\theta(\theta-t)t^2dt = \frac{\theta^4}{12}.$$

Using this inequality, the left side of (4.4) is less than $\sum_1^{n-2}\nu^3\theta_n^3 < n^3\theta_n^3 = O(1)$. This proves the following theorem.

THEOREM 7. *The transform S_n has a limit $\sigma\tau$ (with $\sigma \neq 0$) for every sequence $\{\tau_n\}$ which is summable $(C, 2)$ to τ if and only if $n\theta_n = 2(\lambda\pi + \epsilon_n)$, $\epsilon_n = O(n^{-1/2})$, λ an integer greater than 0. We then have $S_n \to \tau \int_0^{2\lambda\pi} t^{-1} \sin t\, dt = \tau\sigma(2\lambda\pi)$.*

5. **$(C, 2)$ summability of $\{nb_n\}$ for sine series.** We shall prove the following theorem.

THEOREM 8. *If $f(\theta) \sim \sum_1^\infty b_\nu \sin \nu\theta$, and for some j*

$$\int_0^\theta |f(t) - j/2|\, dt = O(\theta), \quad \int_0^\theta \{f(t) - j/2\}\, dt = o(\theta), \text{ as } \theta \downarrow 0, \tag{5.1}$$

then

$$(C, 2) \lim nb_n = j/\pi. \tag{5.2}$$

We have, for $\tau_n = nb_n$,

$$\tau_n^2 = \sum_1^n (n - \nu + 1)\nu b_\nu = \frac{2}{\pi}\int_0^\pi f(t)\left(\sum_1^n (n - \nu + 1)\nu \sin \nu t\right) dt,$$

and $\sum_1^n (n-\nu+1)\nu \sin \nu t = -(d/dt)\sum_1^n (n-\nu+1)\cos \nu t$. On putting

$$\sum_1^n \cos \nu t = -\frac{1}{2} + \frac{\sin (n + 1/2)t}{2 \sin (t/2)} \equiv \sigma_n(t),$$

we have

$$\begin{aligned}\sum_1^n (n - \nu + 1)\cos \nu t = \sum_1^n \sigma_\nu(t) &= -\frac{n}{2} + \frac{1}{2 \sin (t/2)} \sum_1^n \sin (\nu + 1/2)t \\ &= -\frac{n}{2} + \frac{1}{2 \sin (t/2)} \{\sin^2 (n + 1)(t/2) - \sin^2 (t/2)\} \\ &= -\frac{n+1}{2} + \frac{1}{2} \cdot \frac{\sin^2 (n + 1)(t/2)}{\sin^2 (t/2)},\end{aligned}$$

hence

$$\begin{aligned}&\sum_1^n (n-\nu+1)\nu \sin \nu t \\ &= \frac{1}{2} \frac{\sin (n+1)(t/2)}{\sin^3 (t/2)} \left\{\cos \frac{t}{2} \sin (n+1)\frac{t}{2} - (n+1)\cos (n+1)\frac{t}{2} \sin \frac{t}{2}\right\}.\end{aligned} \tag{5.3}$$

Thus

$$(5.4)\qquad \tau_n^2=\frac{1}{\pi}\int_0^\pi f(t)\,\frac{\sin\,(n+1)(t/2)}{\sin^3\,(t/2)}\left\{\sin\,n\,\frac{t}{2}-n\cos\,(n+1)\,\frac{t}{2}\,\sin\,\frac{t}{2}\right\}dt;$$

furthermore

$$\begin{aligned}\int_0^\pi\left(\sum_1^n(n-\nu+1)\nu\sin\nu t\right)dt&\\ =\sum_1^n(n-\nu+1)\{1-(-1)^\nu\}&=2\sum_{\lambda=0}^{(n-1)/2}(n-2\lambda)\\ =4\sum_{\lambda=0}^{(n-1)/2}\left(\frac{n}{2}-\lambda\right)&=h_n,\end{aligned}\qquad(5.5)$$

say, where $2hn/n^2\to 1$ as $n\to\infty$. Our aim is to prove

$$(5.6)\qquad \tau_n^2/h_n\to j/\pi,\qquad\text{or}\qquad(\tau_n^2-jh_n/\pi)/h_n\to 0.$$

From (5.3), (5.4) and (5.5)

$$\tau_n^2-\frac{1}{\pi}jh_n=\frac{1}{\pi}\int_0^\pi\{f(t)-j/2\}\,\frac{\sin\,(n+1)(t/2)}{\sin^3\,(t/2)}\,K_n(t)dt,$$

where

$$\begin{aligned}(5.7)\qquad K_n(t)&=\sin\,(nt/2)-n\cos\,((n+1)t/2)\sin\,(t/2)\\ &=n\sin\,(t/2)(1-\cos\,((n+1)t/2))-(n\sin\,(t/2)-\sin\,(nt/2)).\end{aligned}$$

Now from (3.5)

$$\begin{aligned}0<\frac{1}{\sin^3\,(t/2)}-\frac{1}{(t/2)^3}&=\left(\frac{1}{\sin\,(t/2)}-\frac{2}{t}\right)\left(\frac{1}{\sin^2\,(t/2)}+\frac{2}{t\sin\,(t/2)}+\frac{4}{t^2}\right)\\ &<\frac{t}{6}\cdot\frac{3}{\sin^2\,(t/2)}<\frac{\pi}{2\sin\,(t/2)},\end{aligned}$$

and

$$|K_n(t)|<n\sin\,(t/2)+|\sin\,n(t/2)|.$$

Hence

$$\begin{aligned}&\left|\int_0^\pi\{f(t)-j/2\}\left(\frac{1}{\sin^3\,(t/2)}-\frac{1}{(t/2)^3}\right)\sin\,(n+1)\,\frac{t}{2}\,K_n(t)dt\right|\\ &\qquad<\frac{\pi}{2}\int_0^\pi|f(t)-j/2|\left(n+\frac{|\sin\,(nt/2)|}{\sin\,(t/2)}\right)dt=O(n)=o(h_n);\end{aligned}$$

thus (5.6) will follow if we prove

$$\int_0^\pi \{f(t) - j/2\} \frac{\sin((n+1)t/2)}{t^3} K_n(t)dt = o(n^2).$$

In view of (5.7) this will follow from

$$(5.8)\quad \int_0^\pi \{f(t) - j/2\} \frac{\sin((n+1)t/2)}{t^3} \sin\frac{t}{2}\left(1 - \cos(n+1)\frac{t}{2}\right)dt = o(n),$$

and

$$(5.9)\quad \int_0^\pi \{f(t) - j/2\} \frac{\sin((n+1)t/2)}{t^3}\left(n\sin\frac{t}{2} - \sin n\frac{t}{2}\right)dt = o(n^2).$$

Again, using (3.5),

$$\left|\int_0^\pi \{f(t) - j/2\} \frac{\sin((n+1)t/2)}{t^3}\left(\frac{t}{2} - \sin\frac{t}{2}\right)\left(1 - \cos(n+1)\frac{t}{2}\right)dt\right| < \int_0^\pi |f(t) - j/2|\,dt = O(1),$$

thus (5.8) is equivalent to

$$\int_0^\pi \{f(t) - j/2\} \frac{\sin((n+1)t/2)}{t^2}\left(1 - \cos(n+1)\frac{t}{2}\right)dt = o(n),$$

or

$$I_n \equiv \int_0^\pi \{f(t) - j/2\} \frac{\sin((n+1)t/2) - 2^{-1}\sin(n+1)t}{t^2}dt = o(n).$$

We write $I_n = \int_0^{c/n} + \int_{c/n}^\pi = L_1 + L_2$, say, where c is a constant, arbitrarily large. On putting $\int_0^\theta \{f(t) - j/2\}dt = F(\theta)$, and using integration by parts, we get from (5.1)

$$\begin{aligned} L_1 &= \frac{F(t)}{t}\frac{\sin(n+1)(t/2) - 2^{-1}\sin(n+1)t}{t}\Bigg]_0^{c/n} \\ &\quad - (n+1)\int_0^{c/n} F(t)\frac{\cos((n+1)t/2) - \cos(n+1)t}{2t^2}dt \\ &\quad + 2\int_0^{c/n} F(t)\frac{\sin((n+1)t/2) - 2^{-1}\sin(n+1)t}{t^3}dt \\ &= o(n) + o(n)\cdot\int_0^{c/n} t^{-1}|\sin(3(n+1)t/4)\sin((n+1)t/4)|\,dt \\ &\quad + o(1)\int_0^{c/n} t^{-2}\sin^2((n+1)t/4)dt = o(n), \end{aligned}$$

as $n\to\infty$. Furthermore $|L_2| < \int_{c/n}^{\pi} |f(t)-j/2|\, 2\sin^2((n+1)t/4)/t^2\,dt$; we use here Fejér's inequality, $\sin^2 x < (2x/(1+x))^2$ for $x>0$. Then

$$|L_2| < \frac{(n+1)^2}{2}\int_{c/n}^{\pi} |f(t)-j/2|\,\frac{dt}{(1+((n+1)t/4))^2}$$

$$= \frac{(n+1)^2}{2}\,\frac{\phi(t)}{(1+((n+1)t/4))^2}\Big]_{c/n}^{\pi}$$

$$+\frac{(n+1)^2}{2}\int_{c/n}^{\pi}\frac{(n+1)\phi(t)\,dt}{4(1+((n+1)t/4))^3},$$

where $\phi(\theta)=\int_0^{\theta}|f(t)-j/2|\,dt$. But from (5.1), $\phi(\theta)<\gamma\theta$, γ an absolute constant, hence

$$|L_2| < O(1)+\frac{(n+1)^3}{8}\,\gamma\int_{c/n}^{\pi}\frac{t\,dt}{(1+((n+1)t/4))^3}$$

$$< O(1)+\frac{(n+1)^2}{2}\,\gamma\int_{c/n}^{\pi}\frac{dt}{(1+((n+1)t/4))^2} < \frac{2\gamma(n+1)}{1+c/4}+O(1),$$

and $\limsup_{n\to\infty} n^{-1}L_2 \leqq 8\gamma/(4+c)$. Thus $\limsup_{n\to\infty} n^{-1}|I_n| \leqq 8\gamma/(4+c)$; but c being arbitrary, now (5.8) follows.

Similarly we decompose the left side of (5.9) into $\int_0^{c/n}+\int_{c/n}^{\pi}=M_1+M_2$ say, where

$$M_1=\int_0^{c/n}\{f(t)-j/2\}t^{-3}\left[\frac{n}{2}\left(\cos n\frac{t}{2}-\cos(n+2)\frac{t}{2}\right)\right.$$

$$\left.+\frac{1}{2}\left(\cos(2n+1)\frac{t}{2}-\cos\frac{t}{2}\right)\right]dt$$

$$=\int_0^{c/n}\{f(t)-j/2\}t^{-3}Q_n(t)\,dt$$

$$=F(t)t^{-3}Q_n(t)\Big]_0^{c/n}-\int_0^{c/n}F(t)t^{-3}Q_n'(t)\,dt$$

$$+3\int_0^{c/n}t^{-4}Q_n(t)F(t)\,dt.$$

The mean value theorem yields

$$\frac{Q_n(t)}{n^4t^4}\sim\frac{Q_n'(t)}{4n^4t^3}\sim\frac{Q_n''(t)}{12n^4t^2}\sim\frac{Q'''(t)}{24n^4t}\sim\frac{1}{96}$$

for $t\to 0$, hence, using (5.1), we find that $M_1=o(n^2)$, as $n\to\infty$.

To estimate M_2 we use the formula

$$n \sin x - \sin nx = \sin x \sum_{\nu=1}^{n} \{1 - \cos (n - 2\nu + 1)x\} > 0, \qquad 0 < x < \pi,$$

which follows from

$$\frac{\sin nx}{\sin x} = \frac{e^{inx} - e^{-inx}}{e^{ix} - e^{-ix}} = \sum_{1}^{n} e^{i(n-2\nu+1)x} = \sum_{1}^{n} \cos (n - 2\nu + 1)x.$$

Now

$$|M_2| < \int_{c/n}^{\pi} |f(t) - j/2|\, t^{-3} \sin \frac{t}{2} \sum_{1}^{n} \left\{1 - \cos (n - 2\nu + 1) \frac{t}{2}\right\} dt$$

$$< \int_{c/n}^{\pi} |f(t) - j/2|\, t^{-2} \sum_{2\nu \leqq n} \left\{1 - \cos (n - 2\nu + 1) \frac{t}{2}\right\} dt = \sum_{2\nu \leqq n} D_\nu,$$

say, and, as in the estimate of L_2

$$D_\nu < \phi(\pi) \frac{(n - 2\nu + 1)^2}{2(1 + (n - 2\nu + 1)\pi/4)^2} + \frac{2\gamma(n - 2\nu + 1)}{1 + c/4}$$

$$< \phi(\pi) + \frac{8\gamma(n - 2\nu + 1)}{4 + c}.$$

Hence $|M_2| < n\phi(\pi) + 4\gamma n^2/(4+c)$, and $\limsup_{n\to\infty} |M_2|/n^2 \leqq 4\gamma/(4+c)$; c being arbitrarily large, we finally get (5.9), and Theorem 8 is proved.

6. **The jump of $f(\theta)$ and Gibbs' phenomenon.** The foregoing results enable us to give new formulae for the jump of $f(\theta)$ and to prove a Gibbs phenomenon.

THEOREM 9. *Under the assumptions* (5.1), *j is determined by any one of the formulae*

$$(C, 2) \lim nb_n = \frac{1}{\pi} j, \tag{6.1}$$

$$\sum_{1}^{n} b_\nu \sin \nu \frac{\pi}{n} \to \frac{1}{\pi} j \int_0^{\pi} \frac{\sin t}{t}\, dt, \tag{6.2}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} s_\nu\left(\frac{\pi}{\nu}\right) = \frac{1}{\pi} j \int_0^{\pi} \frac{\sin t}{t}\, dt, \tag{6.3}$$

$$\lim \frac{n}{2\pi} \int_0^{2\pi/n} s_n(t)\, dt = \frac{1}{\pi} j \int_0^{2\pi} \frac{\sin t}{t}\, dt, \tag{6.4}$$

where $s_n(t) = \sum_1^n b_\nu \sin \nu t$.

(6.1) is the statement of Theorem 8; (6.2) follows from Theorem 2 for $\theta_n \equiv 0$, $\phi_n = \pi/n$, $\tau_n = nb_n$, using (6.1). Similarly (6.3) follows from Theorem 4 for $\tau_n = nb_n$, $\theta_n = \pi/n$, and (6.4) follows from Theorem 7 for $\theta_n = 2\pi/n$.

To prove a Gibbs' phenomenon under the same assumptions, put, in Theorem 2, $\tau_n = nb_n$, $\theta_n \equiv 0$, $n\phi_n = \pi + O(1/n)$, then

$$s_n(\phi_n) \to \frac{1}{\pi} j \int_0^\pi t^{-1} \sin t dt,$$

hence

$$\limsup s_n(\theta_n) \geqq j/2 \times 1.08949 \cdots, \qquad \theta_n \downarrow 0.$$

Similarly from Theorem 4 for $\alpha = \pi$

$$\frac{1}{n} \sum_1^n s_\nu(\theta_\nu) \to \frac{1}{\pi} j \int_0^\pi t^{-1} \sin t dt,$$

hence

$$\limsup s_n(\theta_n) \geqq \frac{1}{\pi} j \int_0^\pi t^{-1} \sin t dt.$$

Theorem 7 also proves the presence of a Gibbs' phenomenon, however with a smaller constant.

We may also combine Theorem 8 with Theorem 3 of [3], putting there $\rho_n \equiv 1$, $\kappa = 2$, $n\theta_n = \pi + O(n^{-1})$. Summarizing, we have the following theorem.

THEOREM 10. *Under the assumptions* (5.1)

$$\limsup_{\theta_n \downarrow 0} \sum_1^n b_\nu \sin \nu\theta_n \geqq \frac{jg}{2},$$

where

$$g = \frac{2}{\pi} \int_0^\pi t^{-1} \sin t dt = 1.08949 \cdots.$$

LITERATURE

1. W. Rogosinski, *Über die Abschnitte trigonometrischer Reihen*, Math. Ann. vol. 95 (1926) pp. 110–134.

2. W. Rogosinski, *Über den Einfluss einseitiger Eigenschaften einer Funktion auf ihre Fourierreihe*, Schriften der Königsberger Gelehrten Gesellschaft vol. 3 (1926) pp. 57–98.

3. O. Szász, *On the partial sums of Fourier series at points of discontinuity*, Trans. Amer. Math. Soc. vol. 53 (1943) pp. 440–453.

UNIVERSITY OF CINCINNATI,
CINCINNATI, OHIO.

ON REFLEXIVE NORMS FOR THE DIRECT PRODUCT OF BANACH SPACES

BY

ROBERT SCHATTEN

Introduction. In a previous paper [7](1), for two Banach spaces E_1, E_2, the Banach spaces $E_1 \otimes E_2$, $E_1' \otimes E_2'$, $E_1'' \otimes E_2''$ [7, p. 205] are constructed. If the norm N [7, Definition 3.1] is defined on $E_1 \otimes E_2$, then the associate norm N' [7, Definition 3.2 and Lemma 3.1] is defined on $E_1' \otimes E_2'$. Similarly N'' denotes the norm on $E_1'' \otimes E_2''$.

Among the unsolved problems (mentioned in [7, §6]), are listed the following two:

A. What are the exact conditions imposed upon a crossnorm [7, Definition 3.3] under which $(E_1 \otimes E_2)' = E_1' \otimes E_2'$ holds?

B. Is the associate with every crossnorm also a crossnorm, or do there exist crossnorms whose associates are not crossnorms?

In the present paper we present a "partial" answer to problem A (which we denote by A*), and a "partial" answer to problem B (which we denote by B*).

A*. A uniformly convex crossnorm N sets up the relation $(E_1 \otimes E_2)' = E_1' \otimes E_2'$ if, and only if, $N'' = N$.

B*. For reflexive Banach spaces (that is, such that $E_1'' = E_1$, $E_2'' = E_2$) the associate with every crossnorm is also a crossnorm.

In this paper we also show that the values of a crossnorm for all expressions of rank not greater than 2 do not necessarily determine the crossnorm.

The following should be mentioned in immediate connection with problem A:

It is evident that for norms for which $(E_1 \otimes E_2)' = E_1' \otimes E_2'$ holds, $N'' = N$. Since in general (for any norm N) all we can state is $(E_1 \otimes E_2)' \supset E_1' \otimes E_2'$ [7, p. 205], we have no basis for assuming that N'' represents the norm in $(E_1 \otimes E_2)''$, or $N'' = N$ for expressions in $\mathfrak{A}(E_1, E_2) \subset \mathfrak{A}(E_1'', E_2'')$ [7, Definition 1.3]. Therefore, $N'' \leqq N$ [7, Lemma 3.2] is the best that can be stated in the general case.

In the present paper we present some results on reflexive norms, that is, such that $N'' = N$.

Presented to the Society, April 24, 1943, under the title "On certain properties of crossnorms." Received by the editors March 11, 1943, and, in revised form, June 3, 1943.

(1) Numerals in brackets refer to the bibliography at the end of this paper. Throughout this paper we shall use the results and notation of [7], with a slight modification. Since the printer does not find it easy to handle double and triple bars over expressions by machines, we shall write E' instead of $\overline{E}$, E'' instead of $\overline{\overline{E}}$, E''' instead of $\overline{\overline{\overline{E}}}$. Similarly N' shall take the place of $\overline{N}$, and N'' that of $\overline{\overline{N}}$. Thus $\langle N'' \rangle'$ shall stand for $\overline{\langle \overline{N} \rangle}$.

In §1 we prove that norms that are reflexive, minimal, or have an associate property are identical. In §2 we show that for reflexive Banach spaces the associate with every crossnorm is also a crossnorm, that is, there exists a least crossnorm. In §3 we present a method for construction of reflexive crossnorms and prove certain inequalities. An application of the results of this section to Hilbert spaces permits us to construct semi-self-associate crossnorms, that is, for every natural number k a crossnorm S_k is constructed such that $S_k \neq S$ and $S_k = (S_k)' = S$ for all expressions of rank not greater than k, where S denotes the self-associate crossnorm for Hilbert spaces constructed by F. J. Murray and J. v. Neumann [5]. This last result also proves that a crossnorm is not determined by the values which it assumes for all expressions of rank not greater than 2. Finally, in §4 we show that a uniformly convex crossnorm sets up the relation $(E_1 \otimes E_2)' = E_1' \otimes E_2'$ if, and only if, $N'' = N$.

1. In this section we shall assume that E_1, E_2 are Banach spaces, with no special restrictions.

We introduce the following additional notation:

a. If a norm N is defined on $\mathfrak{A}(E_1, E_2)$, then N'' is defined on $\mathfrak{A}(E_1'', E_2'') \supset \mathfrak{A}(E_1, E_2)$. By $\langle N'' \rangle$ we shall understand N'' considered only on $\mathfrak{A}(E_1, E_2)$. Similarly if N is defined on $\mathfrak{A}(E_1, E_2)$, N''' is defined on $\mathfrak{A}(E_1''', E_2''') \supset \mathfrak{A}(E_1', E_2')$, and $\langle N''' \rangle$ denotes N''' considered only on $\mathfrak{A}(E_1', E_2')$.

b. The set $\mathfrak{A}(E_1, E_2)$ in which there is defined a norm N, we shall denote by $\mathfrak{A}_N(E_1, E_2)$. Thus $\mathfrak{A}_{N'}(E_1', E_2')$ shall denote the set $\mathfrak{A}(E_1', E_2')$ in which there is defined the norm N' associate with N.

c. The symbol $\vdots \bar{f} \in \mathfrak{A}^*(E_1, E_2) \vdots \sup |\overline{F}(\bar{f})|/N(\bar{f})$ shall denote the least upper bound for all numbers $|\overline{F}(\bar{f})|/N(\bar{f})$, obtained when $\bar{f}$ varies over $\mathfrak{A}^*(E_1, E_2)$.

LEMMA 1.1. *If N is a norm in $\mathfrak{A}(E_1, E_2)$, then $N' = \langle N'' \rangle' = N'''$ for $\overline{F}$ in $\mathfrak{A}^*(E_1', E_2') \subset \mathfrak{A}^*(E_1''', E_2''')$.*

Proof. Let $\overline{F} \in \mathfrak{A}^*(E_1', E_2')$. Then

$$N'(\overline{F}) = \vdots \bar{f} \in \mathfrak{A}^*(E_1, E_2) \vdots \sup |\overline{F}(\bar{f})|/N(\bar{f})$$

$$\leqq \vdots \bar{f} \in \mathfrak{A}^*(E_1, E_2) \vdots \sup |\overline{F}(\bar{f})|/N''(\bar{f}) = \langle N'' \rangle'(\overline{F})$$

$$\leqq \vdots \bar{f} \in \mathfrak{A}^*(E_1'', E_2'') \vdots \sup |\overline{F}(\bar{f})|/N''(\bar{f}) = N'''(\overline{F}).$$

On the other hand, $N'''(\overline{F}) \leqq N'(\overline{F})$ [7, Lemma 3.2]. This completes the proof.

DEFINITION 1.1. *A norm N will be termed minimal if, for every norm N^0 for which $N' = (N^0)'$, we have $N^0 \geqq N$.*

DEFINITION 1.2. *A norm N will be termed reflexive if $\langle N'' \rangle = N$.*

DEFINITION 1.3. *A norm N in $\mathfrak{A}(E_1, E_2)$ will be said to have an "associate*

property" if, for a certain norm N^0 in $\mathfrak{A}(E_1', E_2')$, $\mathfrak{A}_N(E_1, E_2) \subset \mathfrak{A}_{(N^0)'}(E_1'', E_2'')$.

THEOREM 1.1. *For a norm N, the following statements are equivalent:*
(a) *N is minimal,*
(b) *N is reflexive,*
(c) *N has an associate property.*

Proof. We shall prove (a)→(b)→(c)→(a). Let N be minimal. By Lemma 1.1, $\langle N'' \rangle' = N'$ for $\overline{F} \in \mathfrak{A}^*(E_1', E_2')$ and any norm N in $\mathfrak{A}(E_1, E_2)$. Thus $\langle N'' \rangle$ and N have the same associate. Since N is minimal by hypothesis, this means $\langle N'' \rangle \geqq N$. On the other hand, for any norm N, $\langle N'' \rangle \leqq N$ [7, Lemma 3.2]. Thus $\langle N'' \rangle = N$. We have proved (a)→(b).

Suppose N is reflexive, that is, $\langle N'' \rangle = N$. Then obviously N has an associate property, because

$$\mathfrak{A}_N(E_1, E_2) = \mathfrak{A}_{\langle N'' \rangle}(E_1, E_2) \subset \mathfrak{A}_{N''}(E_1'', E_2'').$$

Therefore (b)→(c).

Suppose finally that N defined in $\mathfrak{A}(E_1, E_2)$ has an associate property, that is, for a certain norm N^0 defined in $\mathfrak{A}(E_1', E_2')$, $\mathfrak{A}_N(E_1, E_2) \subset \mathfrak{A}_{(N^0)'}(E_1'', E_2'')$. Let the norm N^{00} in $\mathfrak{A}(E_1, E_2)$ be such that $(N^{00})' = N'$. Then, for $\overline{F} \in \mathfrak{A}^*(E_1', E_2')$, we have

$$\begin{aligned}(N^{00})'(\overline{F}) = N'(\overline{F}) &= \vdots \bar{f} \in \mathfrak{A}^*(E_1, E_2) \vdots \sup |\overline{F}(\bar{f})| / N(\bar{f}) \\ &\leqq \vdots \bar{f} \in \mathfrak{A}^*(E_1'', E_2'') \vdots \sup |\overline{F}(\bar{f})| / (N^0)'(\bar{f}) \\ &= (N^0)''(\overline{F}) \leqq N^0(\overline{F}).\end{aligned}$$

Thus, $(N^{00})' \leqq N^0$ throughout $\mathfrak{A}(E_1', E_2')$, and therefore $(N^{00})'' \geqq (N^0)'$ throughout $\mathfrak{A}(E_1'', E_2'')$. Consequently, $N^{00} \geqq (N^{00})'' \geqq (N^0)' = N$, for $\bar{f}$ in $\mathfrak{A}^*(E_1, E_2) \subset \mathfrak{A}^*(E_1'', E_2'')$. Thus N is minimal. Therefore (c)→(a). This completes the proof.

COROLLARY 1. *If N and N^0 denote two reflexive norms, such that $N \leqq N^0$ and $N \neq N^0$, then $N' \geqq (N^0)'$ and $N' \neq (N^0)'$.*

Proof. Obviously $N' \geqq (N^0)'$. If it were $N' = (N^0)'$, then $N'' = (N^0)''$ and $\langle N'' \rangle = \langle (N^0)'' \rangle$. By hypothesis N and N^0 are reflexive. Thus Definition 1.2 gives $N = N^0$. This contradicts our assumption.

COROLLARY 2. *If N and N^0 denote two norms in $\mathfrak{A}(E_1, E_2)$, of which N is reflexive, and $N \leqq N^0$ throughout $\mathfrak{A}(E_1, E_2)$, then $N \leqq \langle (N^0)'' \rangle$.*

Proof. By assumption, $N \leqq N^0$. Therefore $N' \geqq (N^0)'$ and $N'' \leqq (N^0)''$ [7, Lemma 3.3]. Thus $\langle N'' \rangle \leqq \langle (N^0)'' \rangle$. N is reflexive by assumption. Therefore $N = \langle N'' \rangle$, and $N \leqq \langle (N^0)'' \rangle$. This completes the proof.

Remark. Property (c) of Theorem 1.1 suggests the existence of an infinite number of different reflexive crossnorms. We shall prove later that this is

the case, and present a method for construction of reflexive crossnorms(2).

2. Throughout the rest of this paper we shall assume $E_1=E_1''$, $E_2=E_2''$. In this case $\mathfrak{A}(E_1, E_2)=\mathfrak{A}(E_1'', E_2'')$, and for any norm N in $\mathfrak{A}(E_1, E_2)$, $\langle N''\rangle=N''$, $\langle N'''\rangle=N'''$. It should be noticed, however, that many results of the following sections are also valid for general Banach spaces.

Let $N_L(E_1, E_2)$, $N_G(E_1, E_2)$ denote the least crossnorm whose associate is also a crossnorm and the greatest crossnorm in $\mathfrak{A}(E_1, E_2)$, respectively; in the case where there is no fear of misunderstanding, we shall write simply N_L, N_G [7, Definitions 4.1, 4.2, Lemmas 4.1, 4.2 and Theorem 4.1].

LEMMA 2.1. *$N_L(E_1, E_2)=N_L(E_1'', E_2'')$ and $N_G(E_1, E_2)=N_G(E_1'', E_2'')$.*

Proof. This is a consequence of the definition of the least and greatest crossnorms [7, Definitions 4.1, 4.2].

LEMMA 2.2 *For any crossnorm N in $\mathfrak{A}(E_1, E_2)$ for which $N\geqq N_L(E_1, E_2)$, we have $N'\geqq N_L(E_1', E_2')$, $N''\geqq N_L(E_1'', E_2'')=N_L(E_1, E_2)$, $\cdots$.*

Proof. Let N denote a crossnorm whose associate is also a crossnorm, that is, $N\geqq N_L$ [7, Theorem 4.1]. Since $N''\leqq N$ [7, Lemma 3.2], $N''(f\otimes\varphi)\leqq N(f\otimes\varphi)=\|f\|\,\|\varphi\|$ for $f\in E_1$, $\varphi\in E_2$. But N'' is the associate of the crossnorm N'. Therefore $N''(f\otimes\varphi)\geqq\|f\|\,\|\varphi\|$ [7, Lemma 4.3]. Thus N'' is a crossnorm, or the associate with the crossnorm N' in $\mathfrak{A}(E_1', E_2')$ is a crossnorm. This gives $N'\geqq N_L(E_1', E_2')$ [7, Theorem 4.1]. From $N\geqq N_L(E_1, E_2)$ we conclude $N'\geqq N_L(E_1', E_2')$. Similarly from $N'\geqq N_L(E_1', E_2')$ we conclude $N''\geqq N_L(E_1'', E_2'')=N_L(E_1, E_2)$, by Lemma 2.1. This completes the proof.

LEMMA 2.3. *N_L is reflexive.*

Proof. Obviously $N_L\geqq N_L$. Therefore $(N_L)''\geqq N_L$, by virtue of Lemma 2.2. But $(N_L)''\leqq N_L$ [7, Lemma 3.2]. Therefore $(N_L)''=N_L$.

LEMMA 2.4. *The associate with the greatest crossnorm is the least crossnorm, that is, $(N_G(E_1, E_2))'=N_L(E_1', E_2')$.*

Proof. Since $N_G(E_1, E_2)\geqq N_L(E_1, E_2)$, $(N_G(E_1, E_2))'\geqq N_L(E_1', E_2')$ as follows from Lemma 2.2. But $(N_L(E_1', E_2'))'$ is a crossnorm [7, Theorem 4.1]. Thus $(N_L(E_1', E_2'))'\leqq N_G(E_1'', E_2'')=N_G(E_1, E_2)$ [7, Theorem 4.2], or $N_L(E_1', E_2')=(N_L(E_1', E_2'))''\geqq(N_G(E_1, E_2))'$, as follows from Lemma 2.3. Therefore $(N_G(E_1, E_2))'=N_L(E_1', E_2')$. This completes the proof.

COROLLARY. *N_L and N_G' are associate with each other.*

Proof. The associate with N_G' is $N_G'''=N_G'=N_L$, by Lemmas 2.4 and 1.1.

(2) The question of existence of non-reflexive crossnorms is not settled in this paper. It should be noticed that for every non-reflexive norm, $E_1'\otimes E_2'$ would be a proper subset of $(E_1\otimes E_2)'$. In particular it is not settled whether the greatest crossnorm is reflexive. It can be shown, however, that when E_1, E_2 are Hilbert spaces, then the greatest crossnorm is reflexive.

The associate with N_L is $N_L' = N_G''$, by Lemma 2.4.

LEMMA 2.5. *If, for a certain crossnorm N, $N' \geqq N_L$, then $N \geqq N_L$.*

Proof. Lemma 2.2 gives $N'' \geqq N_L$, therefore $N \geqq N'' \geqq N_L$. This completes the proof.

THEOREM 2.1. *Every crossnorm $N \geqq N_L$.*

Proof. Since N is a crossnorm, $N \leqq N_G$ [7, Lemma 4.2]. Therefore $N' \geqq N_G' = N_L$, as follows from Lemma 2.4. Thus Lemma 2.5 gives $N \geqq N_L$. This completes the proof.

THEOREM 2.2. *The associate with every crossnorm is also a crossnorm.*

Proof. This is a consequence of Theorem 2.1 and [7, Theorem 4.1].

THEOREM 2.3. *There exists a least crossnorm.*

Proof. It is obviously N_L, as follows from Theorem 2.1 and [7, Definition 4.2 and Lemma 4.4].

COROLLARY. *For every crossnorm N in $\mathfrak{A}(E_1, E_2)$,*

$$\left\| \sum_{i=1}^{n} F(f_i)\varphi_i \right\| \leqq \|F\| N\left(\sum_{i=1}^{n} f_i \otimes \varphi_i \right) \qquad \text{for } F \in E_1.$$

Proof. This is a consequence of Theorem 2.1 and [7, Theorem 4.1.1].

The question of existence of a norm not less than N_L which is not a crossnorm, and whose associate is a crossnorm, remains open. It is clear that if such a norm exists, it must be non-reflexive.

3. DEFINITION 3.1. *Let E_1, E_2 denote two Banach spaces, and N a crossnorm in $\mathfrak{A}(E_1', E_2')$. We define a sequence of functions $\{N_k\}$ for expressions in $\mathfrak{A}(E_1, E_2) = \mathfrak{A}(E_1'', E_2'')$ in the following way:*

$$N_k\left(\sum_{i=1}^{n} f_i \otimes \varphi_i \right) = \sup \left| \left(\sum_{j=1}^{k} F_j \otimes \phi_j \right)\left(\sum_{i=1}^{n} f_i \otimes \varphi_i \right) \right| \Big/ N\left(\sum_{j=1}^{k} F_j \otimes \phi_j \right)$$

where sup, *that is, the least upper bound, is taken for all sequences of k terms $F_1, \cdots, F_k$ in E_1' and $\phi_1, \cdots, \phi_k$ in E_2'.*

THEOREM 3.1. *For every natural k, N_k is a crossnorm in $\mathfrak{A}(E_1, E_2)$.*

Proof. The proof is similar to that of [7, Theorem 7.2].
From the definition, $N_L = N_1 \leqq N_2 \leqq N_3 \leqq \cdots$ and $\lim_{k\to\infty} N_k = N'$.

THEOREM 3.2. *For every natural k, the crossnorm N_k is reflexive.*

Proof. Suppose that for a certain expression $\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0 \in \mathfrak{A}(E_1, E_2)$, $N_k''\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) < N_k\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) - \epsilon$, where ϵ is a certain positive number.

Then for a given expression $\sum_{j=1}^{m} F_j \otimes \phi_j$ in $\mathfrak{A}(E_1', E_2')$, there exists an expression $\sum_{i=1}^{n} f_i \otimes \varphi_i$, such that

$$\frac{\left|\left(\sum_{j=1}^{m} F_j \otimes \phi_j\right)\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right)\right| N_k\left(\sum_{i=1}^{n} f_i \otimes \varphi_i\right)}{\left|\left(\sum_{j=1}^{m} F_j \otimes \phi_j\right)\left(\sum_{i=1}^{n} f_i \otimes \varphi_i\right)\right|} < N_k\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) - \epsilon.$$

This means

$$\frac{\left|\left(\sum_{j=1}^{m} F_j \otimes \phi_j\right)\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right)\right| \left|\left(\sum_{j=1}^{k} \hat{F}_j \otimes \hat{\phi}_j\right)\left(\sum_{i=1}^{n} f_i \otimes \varphi_i\right)\right|}{\left|\left(\sum_{j=1}^{m} F_j \otimes \phi_j\right)\left(\sum_{i=1}^{n} f_i \otimes \varphi_i\right)\right| N\left(\sum_{j=1}^{k} \hat{F}_j \otimes \hat{\phi}_j\right)} < N_k\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) - \epsilon$$

for every sequence of k terms $\hat{F}_1, \cdots, \hat{F}_k$ in E_1'; $\hat{\phi}_1, \cdots, \hat{\phi}_k$ in E_2'. Taking originally for $\sum_{j=1}^{m} F_j \otimes \phi_j$ an expression of the form $\sum_{j=1}^{k} \hat{F}_j \otimes \hat{\phi}_j$ we get $\left|\left(\sum_{j=1}^{k} \hat{F}_j \otimes \hat{\phi}_j\right)\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right)\right| / N\left(\sum_{j=1}^{k} \hat{F}_j \otimes \hat{\phi}_j\right) < N_k\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) - \epsilon$ for every sequence of k terms $\hat{F}_1, \cdots, \hat{F}_k$ in E_1', $\hat{\phi}_1, \cdots, \hat{\phi}_k$ in E_2'. Therefore $N_k\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) \leqq N_k\left(\sum_{i=1}^{s} f_i^0 \otimes \varphi_i^0\right) - \epsilon$, as follows from Definition 3.1. The last inequality cannot hold. This completes the proof.

THEOREM 3.3. *If N is a reflexive crossnorm, then $(N_k)' = N$ for all expressions of rank not greater than k.*

Proof. Obviously $(N_k)' \leqq N$ for all expressions of rank not greater than k as a consequence of Definition 3.1 for N_k, and that of an associate with a given norm. On the other hand $N_k \leqq N'$ everywhere in $\mathfrak{A}(E_1, E_2)$. Therefore $(N_k)' \geqq N'' = N$. This completes the proof.

THEOREM 3.4. *If N is a reflexive crossnorm, then N_k is the least crossnorm whose associate equals N for all expressions of rank not greater than k.*

Proof. Theorem 3.3 gives $(N_k)' = N$ for all expressions of rank not greater than k. On the other hand it is easy to see that if a crossnorm N^0 satisfies the inequality $N^0\left(\sum_{i=1}^{n} f_i \otimes \varphi_i\right) < N_k\left(\sum_{i=1}^{n} f_i \otimes \varphi_i\right)$ for a certain expression $\sum_{i=1}^{n} f_i \otimes \varphi_i$, then there exists an expression $\sum_{j=1}^{k} F_j \otimes \phi_j$, such that $(N^0)'\left(\sum_{j=1}^{k} F_j \otimes \phi_j\right) > N\left(\sum_{j=1}^{k} F_j \otimes \phi_j\right)$. Thus $(N^0)' = N$, for all expressions of rank not greater than k, implies $N^0 \geqq N_k$. This completes the proof.

It is not difficult to see that the last theorem is a generalization of [7, Theorem 4.1].

In the case E_1 and E_2 are Hilbert spaces, we may assume $\mathfrak{A}(E_1, E_2) = \mathfrak{A}(E_1', E_2')$. In this case, N and N' are defined in $\mathfrak{A}(E_1, E_2)$. For the case of Hilbert spaces, F. J. Murray and J. v. Neumann define in $\mathfrak{A}(E_1, E_2)$ a self-associate crossnorm, which we shall denote by S [5].

THEOREM 3.5. *For every natural p, $S_p = (S_p)' = S$ for all expressions of rank not greater than p.*

Proof. Since $S = S'$, $S = S''$. Thus S is reflexive. Theorem 3.3 gives $(S_p)' = S$ for all expressions of rank not greater than p. Now suppose that for a certain expression $\sum_{i=1}^{n} f_i \otimes \varphi_i$ and a certain crossnorm $N \leqq S$, $N'(\sum_{i=1}^{n} f_i \otimes \varphi_i) = S(\sum_{i=1}^{n} f_i \otimes \varphi_i)$. Since $N \leqq S$, $N' \geqq S' = S$. But $S^2 \leqq NN'$ [7, p. 213]. Therefore $N(\sum_{i=1}^{n} f_i \otimes \varphi_i) = S(\sum_{i=1}^{n} f_i \otimes \varphi_i)$. An application of the last remark to the fact that $S_p \leqq S$ and $(S_p)' = S$, for all expressions of rank not greater than p, gives $S_p = S$ for all expressions of rank not greater than p. Thus $S_p = (S_p)' = S$ for all expressions of rank not greater than p. This completes the proof.

THEOREM 3.6. *For every natural p, $S_p \neq S_{p+1}$.*

Proof. That $S_2 \neq S_3$(3) is shown in the following manner: Consider an expression $f_1 \otimes \varphi_1 + f_2 \otimes \varphi_2 + f_3 \otimes \varphi_3$, where f_1, f_2, f_3, and $\varphi_1, \varphi_2, \varphi_3$ form orthonormal sets. A calculation shows that

$$S_2(f_1 \otimes \varphi_1 + f_2 \otimes \varphi_2 + f_3 \otimes \varphi_3) = \sup \frac{|(h_1 \otimes \chi_1 + h_2 \otimes \chi_2)(f_1 \otimes \varphi_1 + f_2 \otimes \varphi_2 + f_3 \otimes \varphi_3)}{S(h_1 \otimes \chi_1 + h_2 \otimes \chi_2)}$$

where sup, that is, the least upper bound, is taken for all pairs h_1, h_2, in $\mathfrak{M}_1$ (the closed linear manifold determined by f_1, f_2, f_3) and χ_1, χ_2, in $\mathfrak{M}_2$ (the closed linear manifold determined by $\varphi_1, \varphi_2, \varphi_3$). From this it is easily concluded that $S_3(f_1 \otimes \varphi_1 + f_2 \otimes \varphi_2 + f_3 \otimes \varphi_3) \leqq 2^{1/2}$ and

$$S_3(f_1 \otimes \varphi_1 + f_2 \otimes \varphi_2 + f_3 \otimes \varphi_3) = S(f_1 \otimes \varphi_1 + f_2 \otimes \varphi_2 + f_3 \otimes \varphi_3) = 3^{1/2}.$$

A similar reasoning can be applied to prove $S_p \neq S_{p+1}$ for any natural p.

COROLLARY. $S_p \neq S$ for $p = 1, 2, 3, \cdots$.

Proof. Obviously $S_p \leqq S_{p+1} \leqq S$. But $S_p \neq S_{p+1}$. Therefore $S_p \neq S$. This completes the proof(4).

(3) This result and proof is due to F. J. Murray.

(4) For a reflexive (or not) norm N, N_k is reflexive. There exist, however, reflexive norms, for instance S, such that for no reflexive norm N is $S = N_k$. Proof. For a reflexive N, Definition 3.1 and Theorem 3.3 give $((N_k)')_k = N_k$. If it were $S = N_k$ for a reflexive N, then $(N_k)' = S' = S$ and $S_k = S$ as a consequence of the preceding relation. We have shown above that this is not the case. This completes the proof.

Remark. From Theorems 3.5 and 3.6 it is evident that we have constructed three different reflexive crossnorms, which are equal for all expressions of rank not greater than p; namely S_p, $(S_p)'$, S.

S_p and $(S_p)'$ are reflexive, hence associate with each other.

Thus for every natural p we have constructed reflexive "semi-self-associate" crossnorms S_p and $(S_p)'$, that is not self-associate, but equal to their associates for all expressions of rank not greater than p.

Incidentally, this result also proves that the values of a crossnorm for all expressions of rank not greater than p (where p denotes any natural number) do not necessarily determine the crossnorm.

4. In the introduction of this paper, it was pointed out that if a norm sets up the relation $(E_1 \otimes E_2)' = E_1' \otimes E_2'$, then $N'' = N$ for expressions in $\mathfrak{A}(E_1, E_2) \subset \mathfrak{A}(E_1'', E_2'')$, or N is reflexive. In the present section we consider a "converse" problem.

LEMMA 4.1. *If N is a reflexive crossnorm in $\mathfrak{A}(E_1, E_2)$, then $E_1 \otimes E_2 \subset E_1'' \otimes E_2''$.*

Proof. The linear set $\mathfrak{A}^*(E_1'', E_2'')$ in which there is defined the norm N'' is an extension of the linear set $\mathfrak{A}^*(E_1, E_2)$ in which there is defined the norm N. Thus the closure of $\mathfrak{A}^*(E_1'', E_2'')$ is an extension of the closure of $\mathfrak{A}^*(E_1, E_2)$. This completes the proof.

LEMMA 4.2. *If N is a reflexive crossnorm in $\mathfrak{A}(E_1, E_2)$, then $E_1' \otimes E_2'$ forms a fundamental subset of $(E_1 \otimes E_2)'$, that is if, for all $\overline{F}$ in $E_1' \otimes E_2'$ and a certain $\bar{f}_0$ in $E_1 \otimes E_2$, $\overline{F}(\bar{f}_0) = 0$, then $\bar{f}_0 = 0$.*

Proof. For $\overline{F}_0 \in E_1' \otimes E_2'$, we have $N'(\overline{F}_0) = \sup |\overline{F}_0(\bar{f})| / N(\bar{f})$ where sup is taken over the set of all $\bar{f}$'s in $E_1 \otimes E_2$ [7, Lemma 3.4]. Similarly, for $\overline{F}_0^* \in E_1'' \otimes E_2''$, $N''(\overline{F}_0^*) = \sup |\overline{F}_0^*(\overline{F})| / N'(\overline{F})$ where sup is taken over the set of all $\overline{F}$'s in $E_1' \otimes E_2'$. N is reflexive by assumption. Therefore for $\bar{f}_0 \in E_1 \otimes E_2$, Lemma 4.1 gives $N''(\bar{f}_0) = \sup |\overline{F}(\bar{f}_0)| / N'(\overline{F})$, where sup is taken over the set of all $\overline{F}$'s in $E_1' \otimes E_2'$. The second part of our assumption gives $N''(\bar{f}_0) = 0$. But $N'' = N$. Therefore $N(\bar{f}_0) = 0$. Thus $\bar{f}_0 = 0$. This completes the proof.

LEMMA 4.3. *If a reflexive crossnorm sets up the relation $(E_1 \otimes E_2)'' = E_1 \otimes E_2$, then $(E_1 \otimes E_2)' = E_1' \otimes E_2'$.*

Proof. Suppose that $(E_1 \otimes E_2)'' = E_1 \otimes E_2$, and $\overline{F}^*$ is an element of $(E_1 \otimes E_2)'$ which does not belong to $E_1' \otimes E_2'$. From the construction of $E_1' \otimes E_2'$ follows that the set is closed in $(E_1 \otimes E_2)'$. Hence, there exists a linear functional $\mathcal{J}$ on $(E_1 \otimes E_2)'$, such that $\mathcal{J}(\overline{F}^*) = 1$ and $\mathcal{J}(\overline{F}) = 0$, for $\overline{F}$ in $E_1' \otimes E_2'$ [1, p. 57]. Since $(E_1 \otimes E_2)'' = E_1 \otimes E_2$, there exists an $\bar{f}_0$ in $E_1 \otimes E_2$, corresponding to $\mathcal{J}$, such that $\overline{F}^*(\bar{f}_0) = 1$ and $\overline{F}(\bar{f}_0) = 0$ for $\overline{F}$ in $E_1' \otimes E_2'$. The last condition im-

plies $\bar{f}_0=0$ by Lemma 4.2. This contradicts $\bar{F}^*(\bar{f}_0)=1$. This completes the proof.

THEOREM 4.1. *If a reflexive crossnorm N defined on $\mathfrak{A}(E_1, E_2)$ is uniformly convex* [2], *then* $(E_1\otimes E_2)'=E_1'\otimes E_2'$.

Proof. By continuity it follows that N is uniformly convex in $E_1\otimes E_2$, hence [3, 4, 6] $(E_1\otimes E_2)''=E_1\otimes E_2$. An application of Lemma 4.3 completes the proof.

REFERENCES

1. S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.

2. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. vol. 40 (1936) pp. 396–414.

3. M. M. Day, *Reflexive Banach spaces not isomorphic to uniformly convex spaces*, Bull. Amer. Math. Soc. vol. 47 (1941) pp. 313–317.

4. D. Milman, *On some criteria for the regularity of spaces of type* (B), C. R. (Doklady) Acad. Sci. URSS. vol. 20 (1938) pp. 243–246.

5. F. J. Murray and J. v. Neumann, *On rings of operators*, Ann. of Math. (2) vol. 37(1936) pp. 116–229.

6. B. J. Pettis, *Proof that every uniformly convex space is reflexive*, Duke Math. J. vol. 5 (1939) pp. 249–253.

7. R. Schatten, *On the direct product of Banach spaces*, Trans. Amer. Math. Soc. vol. 53 (1943) pp. 195–217.

UNITED STATES ARMY,
ATLANTA, GA.

QUASIGROUPS. I

BY

A. A. ALBERT

1. **Introduction.** A theory of non-associative algebras has been developed[1] without any assumption of a substitute for the associative law, and the basic structure properties of such algebras have been shown to depend upon the possession of almost these same properties by related associative algebras.

It seems natural then to attempt to obtain an analogous treatment of quasigroups. We shall present the results here. Most of the results in the literature on quasigroups do depend upon special associativity conditions[2] but no assumption of such conditions is necessary for our theorems.

Every quasigroup $\mathfrak{G}$ may be associated with the group $\mathfrak{G}_r$ of nonsingular transformations generated by its multiplications. The isotopy of two quasigroups may then be defined and, as in the case of algebras, two groups (that is associative quasigroups) are isotopic if and only if they are isomorphic. Every quasigroup is isotopic to a *loop*, that is, a quasigroup with an identity element, and we derive all further results for loops.

The concepts of *coset* and *normal divisor* may be defined for loops without any assumption of associativity[3]. Then a subloop $\mathfrak{H}$ of $\mathfrak{G}$ is a normal divisor of $\mathfrak{G}$ if and only if $\mathfrak{H}=e\Gamma$ where e is the identity of $\mathfrak{G}$ and Γ is a normal divisor of the group $\mathfrak{G}_r$. Isotopic loops have corresponding normal divisors, and loops which are isotopic to simple loops are simple. We also combine the con-

Presented to the Society, September 13, 1943. Received by the editors June 17, 1943.

(1) See the paper of N. Jacobson, *A note on non-associative algebras*, Duke Math. J. vol. 3 (1937) pp. 544–548, as well as the author's papers *Non-associative algebras. I. Fundamental concepts and isotopy*, Ann. of Math. vol. 43 (1942) pp. 686–707; *The radical of a non-associative algebra*, Bull. Amer. Math. Soc. vol. 48 (1942) pp. 891–897.

(2) Cf. A. Suschkewitsch, *On a generalisation of the associative law*, Trans. Amer. Math. Soc. vol. 31 (1929) pp. 204–214. A suggestion of isotopy theory is given there and the assumption is made that the right multiplications form a group. Then he shows that the quasigroup is isotopic to a group. See also D. C. Murdoch, *Quasigroups which satisfy certain generalized associative laws*, Amer. J. Math. vol. 61 (1939) pp. 509–522; ibid., *Structure of abelian quasigroups*, Trans. Amer. Math. Soc. vol. 49 (1941) pp. 392–409; H. Griffin, *The abelian quasigroups*, Amer. J. Math. vol. 62 (1940) pp. 725–737.

(3) In this connection see B. A. Hausman and O. Ore, *Theory of quasigroups*, Amer. J. Math. vol. 59 (1937) pp. 983–1004. These authors determined associativity conditions in order that the classical coset decompositions be valid. A similar idea was used by D. C. Murdoch, *Note on normality in quasigroups*, Bull. Amer. Math. Soc. vol. 47 (1941) pp. 134–138 but not by G. N. Garrison, in his *Quasigroups*, Ann. of Math. vol. 41 (1940) pp. 474–487, who also discussed the concept of normal divisor but used the finiteness of the system in an essential way. The definition we shall give here for loops seems to be much more natural than these others and reduces to the classical definition in the case of groups.

cepts of isotopy and homorphism to yield a new concept of *homotopy*(4) of loops. Then we show that if a loop $\mathfrak{G}$ is homotopic to a loop $\mathfrak{G}'$ it is homomorphic to a loop which is isotopic to $\mathfrak{G}'$.

2. **Permutation groups.** A *transformation* on a set $\mathfrak{G}$ is defined to be a correspondence

$$S: \qquad g \to gS$$

on $\mathfrak{G}$ to $\mathfrak{G}$. If the correspondence is one-to-one it is usually called a nonsingular correspondence. It will be convenient here to refer to nonsingular transformations as *permutations*. This is standard usage in the case where $\mathfrak{G}$ is a finite set and there seems to be no reason why the terminology should not be extended to the general case.

The set Σ of *all* permutations of $\mathfrak{G}$ forms a group with respect to the operation of product defined by $g(ST)=(gS)T$. It is customary to call this group the *symmetric* group.

If Γ is any set of permutations of $\mathfrak{G}$, and g is any element of $\mathfrak{G}$, we define $g\Gamma$ to be the set of all elements gS for S in Γ. When Γ is a subgroup of Σ it contains the identity transformation I and $g\Gamma$ contains $gI=g$.

We call a subset $\mathfrak{H}$ of $\mathfrak{G}$ a *transitive system* for a permutation group Γ on $\mathfrak{G}$ if hS is in $\mathfrak{H}$ for every h of $\mathfrak{H}$ and S of Γ and if, for every h and k of $\mathfrak{H}$, there is an S in Γ such that $k=hS$. Then we have the following simple lemma:

LEMMA 1. *The transitive systems $\mathfrak{H}$ for a permutation group Γ on $\mathfrak{G}$ are the sets $\mathfrak{H}=h_0\Gamma$ for h_0 any element of $\mathfrak{H}$.*

For if $\mathfrak{H}=h_0\Gamma$ then every element of $\mathfrak{H}$ has the form $h=h_0S$, $(h_0S)T=h_0(ST)$ is in $\mathfrak{H}$. Also $h=h_0S$, $k=h_0T$ implies that $k=(h_0S^{-1})T=h_0(S^{-1}T)$ for $S^{-1}T$ in Γ. Conversely if $\mathfrak{H}$ is a transitive system and h_0 is in $\mathfrak{H}$ there is an S such that $k=h_0S$, $\mathfrak{H}=h_0\Gamma$.

We shall call a permutation group Γ a *transitive or intransitive group* according as $\mathfrak{G}$ itself is or is not a transitive system for Γ. When Γ is transitive we have $g\Gamma=\mathfrak{G}$ for every g of $\mathfrak{G}$. If Γ is intransitive every $g\Gamma$ is a proper subset of $\mathfrak{G}$.

3. **Quasigroups.** A set $\mathfrak{G}$ is said to form a quasigroup if a product

$$a \cdot x$$

is defined to be a unique element of $\mathfrak{G}$ for every a and x of $\mathfrak{G}$ such that the equations

$$(1) \qquad a \cdot x = b, \qquad y \cdot a = b$$

have unique solutions x and y for every a and b of $\mathfrak{G}$. Then every x of $\mathfrak{G}$ determines a transformation

(4) This term has a very different meaning in topology but its use here and in the theory of algebras seems very desirable.

R_x: $$g \to gR_x = g \cdot x$$

of $\mathfrak{G}$. If $aR_x = gR_x$ then $a \cdot x = g \cdot x$ and the second uniqueness hypothesis of (1) implies that $a = g$. Thus every R_x is a permutation of $\mathfrak{G}$. We shall call R_x a *right multiplication* of $\mathfrak{G}$ and shall designate by $\mathfrak{G}_r$ the set of all right multiplications of $\mathfrak{G}$. In general $\mathfrak{G}_r$ is a subset but not necessarily a subgroup of Σ.

In a similar fashion we determine the *left multiplications*

L_x: $$g \to gL_x = x \cdot g$$

of $\mathfrak{G}$. They are also permutations.

A quasigroup now consists of a set $\mathfrak{G}$ of elements g and a corresponding set $\mathfrak{G}_r$ of permutations R_g on $\mathfrak{G}$. The left multiplications L_g are then determined transformations and will be permutations if it is true that $gR_x = gR_y$ only if $x = y$. This requires that the correspondence $g \to R_g$ shall be one-to-one and that if two distinct permutations of $\mathfrak{G}_r$ be applied to the same element of $\mathfrak{G}$ the results shall be distinct. Conversely when these conditions are given $\mathfrak{G}$ will be a quasigroup. For $g \cdot k = gR$, $x = bL_g^{-1}$, $y = bR_g^{-1}$ are all uniquely determined elements of $\mathfrak{G}$ when every R_g and L_g is a nonsingular transformation.

Note that $gS = gT$ is possible for distinct permutations S and T on $\mathfrak{G}$ but that if $xS = xT$ for every x of $\mathfrak{G}$, where S and T are independent of x, then $S = T$. Note also that in the case where $\mathfrak{G}$ is a finite set of n elements $e_1 \cdots e_n$ we are stating that $\mathfrak{G}$ consists of these elements and n permutations $R_1 \cdots R_n$ on them such that if $x_i R_j = x_i R_k$ for *any* i then $j = k$.

4. **Isotopy.** Two quasigroups $\mathfrak{G}$ and $\mathfrak{G}^{(0)}$ are said to have the same order if there is a one-to-one correspondence $g \to g_0$ between them. In our general considerations there will be no loss of generality if we regard all quasigroups of the same order as consisting of the same elements and thus differing only in their right multiplications.

Any quasigroup $\mathfrak{G}$ now consists of a set $\mathfrak{M}$ of elements and a corresponding set $\mathfrak{G}_r$ of right multiplications. Let $\mathfrak{G}^{(0)}$ be a second quasigroup consisting of the same set of elements and a set $\mathfrak{G}_r$ of permutations R_x on $\mathfrak{M}$. Then we shall say that $\mathfrak{G}$ and $\mathfrak{G}^{(0)}$ are isotopic if there exist permutations A, B, C on $\mathfrak{M}$ such that

$$R_x^{(0)} = AR_{xB}C. \tag{2}$$

This is equivalent to the definition of product (g, x) in $\mathfrak{G}^{(0)}$ in terms of that in $\mathfrak{G}$ by

$$(g, x) = (gA \cdot xB)C. \tag{3}$$

Then $(x, g) = (xA \cdot gB)C$ and

$$L_x^{(0)} = BL_{xA}C. \tag{4}$$

Note that if $\mathfrak{G}$ is given then the transformations R_x and L_x are products of

permutations and are permutations. Thus if $\mathfrak{G}$ is a quasigroup and we define $\mathfrak{G}^{(0)}$ by (3) the resulting isotope is a quasigroup. As in the theory of nonassociative algebras the relation of isotopy is an equivalence relation.

5. **Principal isotopes.** Two quasigroups $\mathfrak{G}$ and $\mathfrak{G}^{(0)}$ of the same order are said to be isomorphic if there exists a permutation P on their common set of elements such that $(gP, xP) = (g \cdot x)P$. Then $(gP^{-1} \cdot xP^{-1})P = (a, x)$ and, if we define $A = P^{-1}$, we see that the isomorphism of two quasigroups may be regarded as the instance

$$R_x^{(0)} = AR_{xA}A^{-1}, \qquad L_x^{(0)} = AL_{xA}A^{-1} \tag{5}$$

of isotopy. Moreover we have

LEMMA 2. *Every isotope $\mathfrak{G}^{(1)}$ of a quasigroup $\mathfrak{G}$ is isomorphic to a principal isotope $\mathfrak{G}^{(0)}$ defined by*

$$R_x^{(0)} = AR_{xB}, \qquad L_x^{(0)} = BL_{xA}. \tag{6}$$

For let $R_x^{(1)} = SR_{xT}U$. Then $\mathfrak{G}^{(1)}$ is isomorphic to $\mathfrak{G}^{(0)}$ defined by $R_x^{(0)} = UR_x^{(1)}U^{-1} = AR_{xB}$ where $A = US$, $B = UT$. The second formula of (6) follows from the consequence (4) of (2) in the case $C = I$.

Principal isotopes are more convenient to study than general isotopes and we shall use Lemma 2 frequently in our proofs.

6. **Loops.** A quasigroup $\mathfrak{G}$ is said to have an identity element e if $e \cdot x = x \cdot e = x$ for every x of $\mathfrak{G}$. Then $xR_e = xL_e = x$ for every x, and e is that unique element of $\mathfrak{G}$ such that $R_e = L_e$ is the identity transformation I. It will be very convenient to have a special name for such quasigroups and, as we said in our introduction, we shall call them loops.

We shall show that every quasigroup is isotopic to a loop. A construction of all loops will then yield quasigroups from which all quasigroups may be constructed. Thus it will be at least reasonable to restrict all further study to that of loops.

THEOREM 1. *Let f and g be any elements of a quasigroup $\mathfrak{G}$ so that there exists a unique h in $\mathfrak{G}$ such that $f = g \cdot h$. Then f is the identity element of the loop which is the principal isotope of $\mathfrak{G}$ defined by* (6) *with*

$$A = R_h^{-1}, \qquad B = L_g^{-1}. \tag{7}$$

Conversely if a loop $\mathfrak{G}^{(0)}$ is a principal isotope of $\mathfrak{G}$ it is defined by (6) *and* (7) *for elements g and h of $\mathfrak{G}$ such that $f = g \cdot h$ is the identity element of $\mathfrak{G}^{(0)}$.*

For if $f = g \cdot h = gR_h = hL_g$ we may write $g = fA$, $h = fB$ for A and B of (7). Then $R_f^{(0)} = AR_{fB} = AR_h = I$, $L_f^{(0)} = BL_{fA} = BL_g = I$. Conversely if $\mathfrak{G}^{(0)}$ is defined by (6) and f is the identity element of $\mathfrak{G}^{(0)}$ then $R_f^{(0)} = AR_{fB} = L_f^{(0)} = BL_{fA} = I$. Put $g = fA$ and $h = fB$ to obtain (7) as well as $f = gA^{-1} = gR_h = g \cdot h$.

7. **Groups.** A quasigroup $\mathfrak{G}$ becomes a group when we assume the associative law $x\cdot(a\cdot y)=(x\cdot a)\cdot y$ for every a, x, y of $\mathfrak{G}$. Then $aR_yL_x=aL_xR_y$. Hence $\mathfrak{G}$ is a group if and only if $R_xL_y=L_yR_x$ for all right and left multiplications of $\mathfrak{G}$. It is known that the group property implies the existence of an identity element, that is, an associative quasigroup is always a loop. We also have

LEMMA 3. *A loop $\mathfrak{G}$ is a group if and only if $\mathfrak{G}_r$ is a group. Then $\mathfrak{G}$ and $\mathfrak{G}_r$ are isomorphic.*

For $x\to R_x$ is a one-to-one mapping of $\mathfrak{G}$ on $\mathfrak{G}_r$. If $\mathfrak{G}$ is a group we have $(a\cdot x)y=aR_xR_y=a\cdot(x\cdot y)=aR_{x\cdot y}$, $R_xR_y=R_{x\cdot y}$, our correspondence is an isomorphism, $\mathfrak{G}_r$ is a group. Conversely let $\mathfrak{G}$ be a loop with identity element e so that

$$eR_x = eL_x = x \tag{8}$$

for every x of $\mathfrak{G}$. If $\mathfrak{G}_r$ is a group we have $R_xR_y=R_g$ where g is a uniquely determined element of $\mathfrak{G}$. For the permutations S of $\mathfrak{G}_r$ are in a one-to-one correspondence $S\to g$ with the elements g of $\mathfrak{G}$ such that $S=R_g$. But $eR_x\cdot eR_y=eR_x\cdot y=eR_xR_y=eR_g=g=x\cdot y$, $R_xR_y=R_{x\cdot y}$, $g\to R_g$ is an isomorphism of the group $\mathfrak{G}_r$ and $\mathfrak{G}$.

Note that in the case of an arbitrary loop we have $x\cdot y=eR_x\cdot eR_y=eR_xR_y=g=eR_g$, but we do not have $R_xR_y=R_g$. For R_xR_y may not even be in $\mathfrak{G}_r$.

We may now prove the important theorem:

THEOREM 2. *A loop is isotopic to a group if and only if it is isomorphic to the group and thus is itself a group.*

For let $\mathfrak{G}^{(0)}$ be a loop isotopic to a group $\mathfrak{G}$ so that we may assume $\mathfrak{G}^{(0)}$ to be a principal isotope of $\mathfrak{G}$ defined by (6), (7). Let e be the identity element of $\mathfrak{G}$, so that, if $y=x^{-1}$, $x\cdot y=e$, we have $R_{x\cdot y}=R_xR_y=I$, $R_y=R_x^{-1}$. Similarly $L_{x^{-1}}=(L_x)^{-1}$. Then $xB=xL_g^{-1}=xL_{g^{-1}}=g^{-1}\cdot x$, $xA=xR_h^{-1}=x\cdot h^{-1}$. But $f^{-1}=(g\cdot h)^{-1}=h^{-1}\cdot g^{-1}$ and

$$R_x^{(0)} = R_h^{-1}R_{xB} = R_{f^{-1}\cdot x}. \tag{9}$$

Similarly $L_x^{(0)}=L_g^{-1}L_{xA}=L_{g^{-1}}L_{x\cdot h^{-1}}=L_{x\cdot f^{-1}}$. But then every $R_x^{(0)}$ is commutative with every $L_y^{(0)}$, $\mathfrak{G}^{(0)}$ is associative. By Lemma 3, $\mathfrak{G}$ and $\mathfrak{G}_r$ are isomorphic, $\mathfrak{G}_r^{(0)}=\mathfrak{G}_r$, $\mathfrak{G}$ and $\mathfrak{G}^{(0)}$ are isomorphic.

8. **Subloops.** The right multiplications R_h of the elements h of any subset $\mathfrak{H}$ of a loop $\mathfrak{G}$ generate a group of transformations which we shall designate by $\mathfrak{H}_r$. It consists of all finite products of positive, negative and zero integral powers of the transformations R_h for h in $\mathfrak{H}$. The left multiplications L_h also generate a transformation group $\mathfrak{H}_l$, and the right and left multiplications

together a group $\mathfrak{H}_r$. All of these groups are subgroups of the group $\mathfrak{G}_r$ defined as above for $\mathfrak{H}=\mathfrak{G}$.

A subloop of a loop $\mathfrak{G}$ is a subset $\mathfrak{H}$ which forms a loop with respect to the defining operation of $\mathfrak{G}$. It is sufficient that $\mathfrak{H}$ form a quasigroup, that is, that $\mathfrak{H}$ contain all products $h\cdot k$ and the solutions x and y of $x\cdot h=k$, $h\cdot y=k$ taken when h and k are in $\mathfrak{H}$. For then $\mathfrak{H}$ will contain the solution e of $x\cdot h=h$. Thus *the identity element of a loop $\mathfrak{G}$ is the identity element of all of its subloops.*

LEMMA 4. *A set $\mathfrak{H}$ is a subloop of $\mathfrak{G}$ if and only if $e\mathfrak{H}_r=\mathfrak{H}$.*

For let $\mathfrak{H}$ be a subloop of $\mathfrak{G}$ so that $\mathfrak{H}$ contains $eI=e$. The elements of $\mathfrak{H}_r$ are products of transformations R_h, L_h, R_h^{-1}, L_h^{-1}, and if $k=eS$ is in $\mathfrak{H}$, then $e(SR_h)=k\cdot h$, $e(SL_h)=h\cdot k$, $x=(eS)R_h^{-1}=kR_h^{-1}$, $y=(eS)L_h^{-1}=kL_h^{-1}$ are all in $\mathfrak{H}$. Hence $e\mathfrak{H}_r$ is contained in $\mathfrak{H}$, every $h=eR_h$ is contained in $e\mathfrak{H}_r$, $e\mathfrak{H}_r=\mathfrak{H}$. Conversely if $e\mathfrak{H}_r=\mathfrak{H}$ we have $eR_hR_k=h\cdot k$ and the solutions $x=eR_kR_k^{-1}$, $y=eL_kL_k^{-1}$ of $x\cdot h=k$ and $h\cdot y=k$ in $\mathfrak{H}$. Then $\mathfrak{H}$ is a subloop.

It should be observed that the order of a subloop $\mathfrak{H}$ of a finite loop $\mathfrak{G}$ need not divide the order of $\mathfrak{G}$. For example let $\mathfrak{G}$ consist of $e=e_1, e_2, e_3, e_4, e_5$, and let $R_{e_i}=R_i$ be defined by $R_e=I$, $R_2=(12)(345)$, $R_3=(13)(254)$, $R_4=(14)(235)$, $R_5=(15)(243)$. Evidently $e_i\cdot e=e_iR_e=e_i$, $e\cdot e_i=eR_i=e_i$, e is the identity element of $\mathfrak{G}$. We compute $e_2\cdot e_2=e_1$, $e_2\cdot e_3=e_5$, $e_2\cdot e_4=e_3$, $e_2\cdot e_5=e_4$ so that $L_2(12)(354)$. By symmetry $L_3=(13)(245)$, $L_4=(14)(253)$, $L_5=(15)(234)$ and $\mathfrak{G}$ is a loop. Its four subloops (e, e_i), defined for $i=2, 3, 4, 5$, are all groups of order two.

9. **Cosets.** Every subgroup Γ of $\mathfrak{G}_r$ defines subsets $x\Gamma$ of $\mathfrak{G}$ which we shall call cosets of $\mathfrak{G}$ relative to Γ. They are transitive systems of Γ by Lemma 1. Thus every element of $\mathfrak{G}$ is in one and only one coset and the sets $x\Gamma$ provide a coset decomposition of $\mathfrak{G}$.

If $\mathfrak{H}$ is a subloop of a loop $\mathfrak{G}$ we shall call the cosets $x\mathfrak{H}_\rho$ the right-hand cosets of $\mathfrak{G}$ relative to $\mathfrak{H}$. Similarly, we shall call the cosets $x\mathfrak{H}_\lambda$ the left-hand cosets of $\mathfrak{G}$ relative to $\mathfrak{H}$, and the cosets $x\mathfrak{H}_r$ the two-sided cosets of $\mathfrak{G}$ relative to $\mathfrak{H}$.

The sets $x\mathfrak{H}$ consisting of all elements $x\cdot h$ for h in the subloop $\mathfrak{H}$ do not, in general, have the property that each element is in one and only one such set. For if $x\cdot h=y\cdot k$ then $y=xR_hR_k^{-1}$ is in $x\mathfrak{H}_\rho$ but may not be in $x\mathfrak{H}$. We shall not call such sets cosets. However, our definition of normal divisors will have the consequence $x\mathfrak{H}_\rho=x\mathfrak{H}_r=x\mathfrak{H}$.

With every subgroup Γ of $\mathfrak{G}$ and corresponding coset decomposition of $\mathfrak{G}$ we associate the set Γ^* of all transformations S of $\mathfrak{G}_r$ such that xS is in $x\Gamma$ for every x of $\mathfrak{G}$. Then $x\Gamma^*$ is contained in $x\Gamma$. But clearly Γ is contained in Γ^*, $x\Gamma$ is in $x\Gamma^*$, $x\Gamma=x\Gamma^*$ for every x. Also we have

LEMMA 5. *The sets Γ^* are subgroups of $\mathfrak{G}_r$.*

For if H and K are in Γ^* and x is in $\mathfrak{G}$ we have $y=xH=xS$ where S is

in Γ. Then $yK=yT$ where T is in Γ, $(xH)K=(xS)T$, $x(HK)=x(ST)=xU$ for $U=ST$ in Γ, HK is in Γ^*. Also $z=xH^{-1}$ has the property $zH=zS$ where S is in Γ if H is in Γ^*, $x=zS$, $x\Gamma=zS\Gamma=z\Gamma$ since Γ is a group, $S\Gamma=\Gamma$ if S is in Γ. Then $z=xH^{-1}$ is in $x\Gamma$ and H^{-1} is in Γ^*.

Note that every subloop $\mathfrak{H}$ of $\mathfrak{G}$ determines a unique subgroup $\mathfrak{H}_\rho^*$ of $\mathfrak{G}_r$ such that $eH_\rho^*=e\mathfrak{H}_\rho=\mathfrak{H}=e\mathfrak{H}_r$ by Lemma 3. However $\mathfrak{H}_r$ is not necessarily a subgroup of $\mathfrak{H}_\rho^*$.

10. **Normal divisors.** A homomorphism of a loop $\mathfrak{G}$ on a loop $\mathfrak{G}'$ is a mapping $g\to g'$ of $\mathfrak{G}$ on $\mathfrak{G}'$ such that every element of $\mathfrak{G}'$ is the image g' of an element g of $\mathfrak{G}$, and $(x\cdot y)'=x'\cdot y'$ for every x and y of $\mathfrak{G}$. The mapping $x\to x\mathfrak{H}_\rho$ of $\mathfrak{G}$ on the quotient set $\mathfrak{G}/\mathfrak{H}$ of all cosets $x\mathfrak{H}_\rho$ satisfies the first of these properties. We shall call a subloop $\mathfrak{H}$ of $\mathfrak{G}$ a *normal divisor* of $\mathfrak{G}$ if $\mathfrak{G}/\mathfrak{H}$ is a loop and $x\to x\mathfrak{H}_\rho$ is a homomorphism of $\mathfrak{G}$ on $\mathfrak{G}/\mathfrak{H}$.

The hypothesis that $\mathfrak{G}/\mathfrak{H}$ is a loop implies that either of the equations

$$x\mathfrak{H}_\rho\cdot y\mathfrak{H}_\rho = x\mathfrak{H}_\rho\cdot w\mathfrak{H}_\rho, \qquad y\mathfrak{H}_\rho\cdot x\mathfrak{H}_\rho = w\mathfrak{H}_\rho\cdot x\mathfrak{H}_\rho \tag{10}$$

is true if and only if $y\mathfrak{H}_\rho=w\mathfrak{H}_\rho$. The homomorphism hypothesis implies that

$$x\mathfrak{H}_\rho\cdot y\mathfrak{H}_\rho = (x\cdot y)\mathfrak{H}_\rho, \tag{11}$$

and thus defines the product operation in $\mathfrak{G}/\mathfrak{H}$. Conversely (11) and (10) imply that $\mathfrak{G}/\mathfrak{H}$ is a loop and that $x\to x\mathfrak{H}_\rho$ is a homomorphism of $\mathfrak{G}$ on $\mathfrak{G}/\mathfrak{H}$. For $\mathfrak{H}=e\mathfrak{H}_\rho$ and by (11), we have $\mathfrak{H}\cdot y\mathfrak{H}_\rho=y\mathfrak{H}_\rho$, $x\mathfrak{H}_\rho\cdot\mathfrak{H}=x\mathfrak{H}_\rho$, $\mathfrak{H}$ is the identity element of $\mathfrak{G}/\mathfrak{H}$. Also $g\mathfrak{H}_\rho\cdot y\mathfrak{H}_\rho=(g\cdot y)\mathfrak{H}_\rho=b\mathfrak{H}_\rho$ and $x\mathfrak{H}_\rho\cdot g\mathfrak{H}_\rho=(x\cdot g)\mathfrak{H}_\rho=b\mathfrak{H}_\rho$ are satisfied by $y=bL_g^{-1}$, $x=bR_g^{-1}$ and the corresponding coset solutions are unique by (10). Then $\mathfrak{G}/\mathfrak{H}$ is a loop as desired.

Note that if

$$\Gamma=\mathfrak{H}_\rho^*, \qquad \mathfrak{H}=e\Gamma, \qquad x\mathfrak{H}_\rho=x\Gamma \tag{12}$$

for every x, and (11) becomes

$$x\Gamma\cdot y\Gamma = (x\cdot y)\Gamma. \tag{13}$$

Also we may use (11) to write (10) in the form

$$(x\cdot y)\Gamma=(w\cdot y)\Gamma, \qquad (y\cdot x)\Gamma=(y\cdot w)\Gamma \tag{14}$$

if and only if $x\Gamma=w\Gamma$. We are now in a position to prove the fundamental theorem:

THEOREM 3. *A subset $\mathfrak{H}$ of a loop $\mathfrak{G}$ with identity element e is a normal divisor of $\mathfrak{G}$ if and only if $\mathfrak{H}=e\Gamma$ where Γ is a normal divisor of $\mathfrak{G}_r$. Then $\Gamma^*=\mathfrak{H}_\rho^*$ is also a normal divisor of $\mathfrak{G}_r$ and*

$$x\Gamma=x\mathfrak{H}_\rho=x\mathfrak{H}=x\mathfrak{H}_\lambda=x\mathfrak{H}_r \tag{15}$$

for every x.

For let $\mathfrak{H}$ be a normal divisor of $\mathfrak{G}$ and define $\Gamma=\mathfrak{H}_\rho^*$ so that we have (12), (13), (14). Since every permutation of $\mathfrak{G}$ is a finite product of permutations R_g, R_g^{-1}, L_g, L_g^{-1} the group Γ will be a normal divisor of $\mathfrak{G}_r$ if we can show that $T^{-1}\Gamma T$ is contained in Γ for $T=R_g$, R_g^{-1}, L_g, L_g^{-1}. Now for every x and y of $\mathfrak{G}$ and S of Γ we define $w=xR_ySR_y^{-1}$ so that $xR_yS=wR_y$, $(x\cdot y)\Gamma=(w\cdot y)\Gamma$. By (14), $x\Gamma=w\Gamma$, w is in $x\Gamma$. But our definition of $\mathfrak{H}_\rho^*$ implies that $R_ySR_y^{-1}$ is in Γ, $R_y\Gamma R_y^{-1}$ is contained in Γ. Similarly define $w=xL_ySL_y^{-1}$ and obtain $(y\cdot x)S=y\cdot w$, $(y\cdot x)\Gamma=(y\cdot w)\Gamma$, $w\Gamma=x\Gamma$, w is in $x\Gamma$ and $L_y\Gamma L_y^{-1}$ is contained in Γ. We next put $z=xR_y^{-1}$ and have $z\cdot y=x$, $zS\cdot y=(z\cdot y)U$ where U is in Γ by (13). Then $xR_y^{-1}SR_y=xU$ is in $x\Gamma$ and $R_y^{-1}\Gamma R_y$ is contained in Γ. A similar argument with $z=yL_z^{-1}$ yields $x\cdot zS=(x\cdot z)U$, $yL_z^{-1}SL_z=yU$ and implies that $L_z^{-1}\Gamma L_z$ is contained in Γ, $\Gamma=\mathfrak{H}_\rho^*$ is a normal divisor of $\mathfrak{G}_r$.

Conversely let $\mathfrak{H}=e\Gamma$ where Γ is a normal divisor of $\mathfrak{G}_r$. If S and T are in Γ then $eS\cdot eT=eSR_{eT}=eR_{eT}U=eTU$ where U is in Γ, TU is in Γ. Also, $eS\cdot y=eT$ implies that $eSR_y=eR_yU=eT$, $y=eTU^{-1}$ is a unique element of $\mathfrak{H}$. Similarly $x\cdot eS=eSL_x=eL_xU=xU=eT$ has the unique solution $x=eTU^{-1}$ in $\mathfrak{H}$. Thus $\mathfrak{H}$ is a subloop of $\mathfrak{G}$. Now $xS\cdot yT=xSR_{yT}=xR_{yT}U=(x\cdot yT)U=yTR_xU=yL_xVU=(x\cdot y)VU$ where V and U are in Γ. Hence $x\Gamma\cdot y\Gamma$ is contained in $(x\cdot y)\Gamma$. But if S is in Γ we may determine $U=R_ySR_y^{-1}$ and have $(x\cdot y)S=xR_yS=xUR_y=xU\cdot y$, $(x\cdot y)\Gamma$ is contained in $x\Gamma\cdot y\Gamma$ and we have (13). If $(x\cdot y)\Gamma=(x\cdot w)\Gamma$ then $yL_x\Gamma=wL_x\Gamma$. But $L_x\Gamma=\Gamma L_x$ and $y\Gamma=w\Gamma$ as desired. Similarly $(y\cdot x)\Gamma=(w\cdot x)\Gamma$ and $yR_x\Gamma=wR_x\Gamma$, $y\Gamma=w\Gamma$. Now $xS=eL_xS=eUL_x=hL_x=x\cdot h$ where $h=eU$ is in $\mathfrak{H}$ and so $x\Gamma$ is contained in $x\mathfrak{H}$. Also $x\cdot h=x\cdot eS=eSL_x=eL_xU=xU$ for U in Γ, $x\mathfrak{H}=x\Gamma$. It follows that Γ^* contains every R_h. Similarly $h\cdot x=eSR_x=eR_xU=xU=xL_h$, Γ^* contains every L_h, Γ^* contains $\mathfrak{H}_r$, $x\Gamma^*=x\Gamma$ contains $x\mathfrak{H}_r$ which contains $x\mathfrak{H}=x\Gamma$. This gives (15). But $x\mathfrak{H}_\rho=x\Gamma$ and $\Gamma^*=\mathfrak{H}_\rho^*$, $x\Gamma=x\mathfrak{H}_\rho^*$ and we have (13) and (14) for Γ replaced by $\mathfrak{H}_\rho^*$. This proves our theorem.

It should be observed that there may be several normal divisors Γ of $\mathfrak{G}_r$ such that $\mathfrak{H}=e\Gamma$ is the same normal divisor of $\mathfrak{G}$. Indeed $\mathfrak{H}=e\Gamma=e\Delta$ if and only if $\mathfrak{H}$ is a transitive system of both Γ and Δ.

As an expected consequence of our definitions we now have

THEOREM 4. *Let $g\to g'$ be a homomorphism of a loop $\mathfrak{G}$ on a loop $\mathfrak{G}'$ and $\mathfrak{H}$ be the set of all elements h of $\mathfrak{G}$ such that h' is the identity element e' of $\mathfrak{G}'$. Then $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$ and the correspondence $x\mathfrak{H}\to x'$ is an isomorphism of the quotient loop $\mathfrak{G}/\mathfrak{H}$ and $\mathfrak{G}'$.*

For if h and k are in $\mathfrak{H}$ we have $(h\cdot k)'=h'\cdot k'=e'$ and $h\cdot k$ is in $\mathfrak{H}$. Also $x\cdot h=k$ implies that $x'\cdot h'=k'$, $x'=e'$, x is in $\mathfrak{H}$. Similarly $h\cdot y=k$ implies that y is in $\mathfrak{H}$ and that $\mathfrak{H}$ is a subloop of $\mathfrak{G}$. The elements y of $x\mathfrak{H}_r$ are products of x by a finite number of elements h of $\mathfrak{H}$ with some order and grouping. Since every $h'=e'$ we have $y'=x'$ no matter what the order and grouping.

Conversely if $y'=x'$ we may write $y=x\cdot z$ for z in $\mathfrak{G}$ and $y'=x'\cdot z'=x'$. But $\mathfrak{G}'$ is a loop and $z'=e'$, z is in $\mathfrak{H}$, y is in $x\mathfrak{H}$. It follows that $x\mathfrak{H}_\tau \leqq x\mathfrak{H}$ and, since $x\mathfrak{H}_\tau \geqq x\mathfrak{H}_\rho \geqq x\mathfrak{H}$, then

$$x\mathfrak{H}_\tau = x\mathfrak{H}_\rho = x\mathfrak{H}_\lambda = x\mathfrak{H}.$$

Moreover $x'=y'$ implies that $x\mathfrak{H}_\rho = y\mathfrak{H}_\rho$ so that the correspondence $x\mathfrak{H}\to x'$ is one-to-one. We compute $[(x\cdot h)\cdot(y\cdot k)]'=(x\cdot h)'\cdot(y\cdot k)'=x'\cdot y'=(x\cdot y)'$ for every h and k of $\mathfrak{H}$ and see that $x\mathfrak{H}\cdot y\mathfrak{H}=(x\cdot y)\mathfrak{H}$. Conversely $(x\cdot y)\cdot h=x\cdot z$ has a solution z in $\mathfrak{G}$ which may be written in the form $z=y\cdot w$ for w in $\mathfrak{G}$. Then $[x\cdot(y\cdot w)]'=x'\cdot(y\cdot w)'=x'\cdot(y'\cdot w')=[(x\cdot y)\cdot h]'=x'\cdot y'$, $y'=y'\cdot w'$, $w'=e'$, w is in $\mathfrak{H}$, $(x\cdot y)\mathfrak{H}=x\mathfrak{H}\cdot y\mathfrak{H}$. This proves that $x\mathfrak{H}\to x'$ is an isomorphism of $\mathfrak{G}/\mathfrak{H}$ and the loop $\mathfrak{G}'$. Then $\mathfrak{G}/\mathfrak{H}$ is a loop and $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$.

11. **The multiplications of quotient loops.** If $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$ and $\Gamma=\mathfrak{H}_\rho^*$, the elements of $\mathfrak{G}/\mathfrak{H}$ are the cosets $x\Gamma$. The elements of $\mathfrak{G}_\tau/\Gamma$ are the cosets $S\Gamma=\Gamma S$ for S in $\mathfrak{G}_\tau$. Each such coset defines a transformation

$$g\Gamma \to (g\Gamma)(S\Gamma) = (gS)\Gamma. \tag{16}$$

But if $aS\Gamma=gS\Gamma$ we have $a\Gamma S=g\Gamma S$, $a\Gamma=g\Gamma$. Hence (16) is a permutation of $\mathfrak{G}/\mathfrak{H}$. We may identify the permutation (16) with $S\Gamma$ and have shown that the elements of the quotient group $\mathfrak{G}_\tau/\Gamma$ may be regarded as being permutations of the quotient loop $\mathfrak{G}/\mathfrak{H}$.

The right multiplications of $\mathfrak{G}/\mathfrak{H}$ are the permutations

$$g\Gamma \to g\Gamma\cdot x\Gamma = (g\Gamma)(R_x\Gamma),$$

that is, the permutations $R_x\Gamma$ of $\mathfrak{G}_\tau/\Gamma$. Similarly the left multiplications of $\mathfrak{G}/\mathfrak{H}$ are the permutations $L_x\Gamma$. The group $\mathfrak{G}_\tau$ is generated by the right and left multiplications and thus $\mathfrak{G}_\tau/\Gamma$ is generated by the $R_x\Gamma$ and $L_y\Gamma$. But these permutations also generate the group of permutations for $\mathfrak{G}/\mathfrak{H}$.

THEOREM 5. *Let $\mathfrak{H}$ be a normal divisor of $\mathfrak{G}$ so that $\Gamma=\mathfrak{H}_\rho^*$ is a normal divisor of $\mathfrak{G}_\tau$. Then $(\mathfrak{G}/\mathfrak{H})_\tau = \mathfrak{G}_\tau/\Gamma$.*

12. **Finite loops.** In our proof of Theorem 3 we showed that the hypothesis $x\Gamma\cdot y\Gamma=(x\cdot y)\Gamma$ implies that $S^{-1}\Gamma S$ is contained in Γ for every S which is either a right multiplication or a left multiplication. In the case where $\mathfrak{G}$ is finite so is Γ and a comparison of orders then yields $S^{-1}\Gamma S=\Gamma$, $\Gamma=S\Gamma S^{-1}$. Indeed it is true that in the finite case if we have $S^{-1}\Gamma S\leqq\Gamma$, for elements S which generate $\mathfrak{G}_\tau$, the group Γ will be a normal divisor of $\mathfrak{G}_\tau$. Thus we have

THEOREM 6. *Let $\mathfrak{H}$ be a subloop of a finite loop $\mathfrak{G}$ and define $x'=x\mathfrak{H}_\rho$ for every x of $\mathfrak{G}$. Then if $(x\cdot y)'=x'\cdot y'$ for every x and y of $\mathfrak{G}$ the quotient set $\mathfrak{G}/\mathfrak{H}$ is a loop and $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$.*

We may also prove

THEOREM 7. *The order of a normal divisor $\mathfrak{H}$ of a finite loop $\mathfrak{G}$ divides the order of $\mathfrak{G}$.*

For if $\mathfrak{H}$ is a normal divisor of $\mathfrak{G}$ we have seen that every element x of $\mathfrak{G}$ is in one and only one set $x\mathfrak{H}$. If $x\cdot h=x\cdot k$ then $h=k$ so that each coset $x\mathfrak{H}$ has exactly as many elements as does $\mathfrak{H}$. Then the order of $\mathfrak{G}$ is the product of the order of $\mathfrak{H}$ and the order of $\mathfrak{G}/\mathfrak{H}$.

This result evidently implies that if a finite loop $\mathfrak{G}$ is homomorphic to a loop $\mathfrak{G}'$ the order of $\mathfrak{G}'$ divides the order of $\mathfrak{G}$.

13. **Simple loops.** A subgroup Γ of $\mathfrak{G}_r$ defines a proper subset $e\Gamma$ of $\mathfrak{G}$ if and only if Γ is intransitive. As in the theory of groups we call a loop simple if its only proper normal divisor is the identity group. Then we have

THEOREM 8. *A loop $\mathfrak{G}$ is simple if and only if the group $\mathfrak{G}_r$ has no intransitive normal divisor except the identity group.*

Every isotope of a loop $\mathfrak{G}$ is isomorphic to a principal isotope $\mathfrak{G}^{(0)}$ of $\mathfrak{G}$. If the original isotope is a loop so is $\mathfrak{G}^{(0)}$ and we may let f be the identity element of $\mathfrak{G}^{(0)}$. Theorem 1 implies that $\mathfrak{G}_r^{(0)}\leqq\mathfrak{G}_r$. The relation of principal isotopy is symmetric and thus $\mathfrak{G}_r=\mathfrak{G}_r^{(0)}$. We pass to a loop $\mathfrak{G}^{(1)}$ isomorphic to $\mathfrak{G}^{(0)}$ and defined by

$$R_x^{(1)}=CR_{xc}^{(0)}C^{-1},\qquad L_x=CL_{xc}^{(0)}C^{-1}$$

with $C=R_f^{-1}$. Then $\mathfrak{G}^{(1)}$ has e as its identity element. Moreover $\mathfrak{G}_r^{(1)}=CG_r^{(0)}C^{-1}=\mathfrak{G}_r^{(0)}=\mathfrak{G}_r$ since C is in $\mathfrak{G}_r$. The application of Theorem 3 yields

THEOREM 9. *Every loop isotopic to a loop $\mathfrak{G}$ is isomorphic to a loop having precisely the same normal divisors as $\mathfrak{G}$.*

Note that we are not stating that the normal divisors $e\Gamma$ of $\mathfrak{G}^{(1)}$ are isomorphic or even isotopic to the normal divisors of $\mathfrak{G}$. They merely consist of the same elements. If $\mathfrak{H}=e\Gamma$ then multiplication for $\mathfrak{H}$ in $\mathfrak{G}$ is defined by $y\cdot h=yR_h$ and in $\mathfrak{G}^{(1)}$ by $(y,h)=yCAR_{hBC}C^{-1}$. It does not appear necessary for $\mathfrak{H}$ in $\mathfrak{G}$ and in $\mathfrak{G}_r$ to be isotopic.

As a consequence of Theorem 9 we have

THEOREM 10. *All loops isotopic to a simple loop are simple.*

14. **Centers.** The *center* $\mathfrak{C}$ of a loop $\mathfrak{G}$ is the set of all quantities c of $\mathfrak{G}$ such that the commutative and associative laws for products hold whenever c is one of the factors. The condition $x\cdot(y\cdot c)=(x\cdot y)\cdot c$ is equivalent to $R_cL_x=L_xR_c$ for every x, and $c\cdot(x\cdot y)=(c\cdot x)\cdot y$ yields $L_cR_y=R_yL_c$ for every c. But $x\cdot c=c\cdot x$ is equivalent to $L_c=R_c$. Hence every R_c is in the center Λ of the group $\mathfrak{G}_r$, every $c=eR_r$, $\mathfrak{C}=e\Lambda$.

Suppose now that H is in Λ and that $c=eH$. Then $x\cdot c=cL_x=eHL_x=eL_xH=xH$, $H=R_c$. Similarly $c\cdot x=cR_x=eHR_x=eR_xH=xH$, $L_c=R_c$. We also have

$R_cL_z=L_zR_c$ so that $x\cdot(y\cdot c)=(x\cdot y)\cdot c$ and similarly $c\cdot(x\cdot y)=(c\cdot x)\cdot y$. But then $R_{y\cdot c}=R_yR_c=R_cR_y=R_{c\cdot y}$ and $(x\cdot c)\cdot y=x\cdot(c\cdot y)$, $c=eH$ is in $\mathfrak{C}$. Hence $\mathfrak{C}=e\Lambda$.

The correspondence $c\to H=R_c$ is one-to-one since if H is given then $c=eH$. Also $eH\cdot eK=eHR_{eK}=eHK$ for every H and K of Λ. Thus we have

THEOREM 11. *The center $\mathfrak{C}$ of a loop $\mathfrak{G}$ with identity element e is isomorphic to the center Λ of $\mathfrak{G}_r$, and $\mathfrak{C}=e\Lambda$.*

If $\mathfrak{G}$ and $\mathfrak{G}^{(0)}$ are loops which are principal isotopes the group $\mathfrak{G}_r=\mathfrak{G}_r^{(0)}$ and the centers of $\mathfrak{G}$ and $\mathfrak{G}^{(0)}$ are isomorphic. By Lemma 2 we have

THEOREM 12. *Isotopic loops have isomorphic centers.*

15. **Homotopy.** While it has been convenient to define isotopy for algebraic systems consisting of the same elements it is not, of course, necessary. Thus we may define two multiplicative algebraic systems $\mathfrak{G}$ and $\mathfrak{G}'$ to be isotopic if there exist nonsingular mappings A, B, C of $\mathfrak{G}$ on $\mathfrak{G}'$ such that

$$xA\cdot yB=(x\cdot y)C. \tag{17}$$

The concept of homomorphism may now be generalized. We call two mappings A and B of a set $\mathfrak{G}$ on a set $\mathfrak{G}'=\mathfrak{G}A=\mathfrak{G}B$ *equivalent* if there exists a nonsingular transformation N of $\mathfrak{G}$ and a nonsingular transformation S of $\mathfrak{G}'$ such that

$$B=NAS.$$

Then we shall say that a multiplicative system $\mathfrak{G}$ is *homotopic* to a multiplicative system $\mathfrak{G}'$ if there exist equivalent mappings A, B, C on $\mathfrak{G}$ to $\mathfrak{G}'$ such that $\mathfrak{G}'=\mathfrak{G}C$ and (17) holds for all x and y of $\mathfrak{G}$. Then $A=NCS$, $B=QCT$, $xNCS\cdot yQCT=(x\cdot y)C$. Define $(x',y')=x'S\cdot y'T$ for every x' and y' of $\mathfrak{G}'$ and $(x,y)=xN^{-1}\cdot yQ^{-1}$ to obtain $(xC,yC)=(x,y)C$. The argument is readily reversed and we see that *$\mathfrak{G}$ is homotopic to $\mathfrak{G}'$ if and only if an isotope of $\mathfrak{G}$ is homomorphic to an isotope of $\mathfrak{G}'$.*

We note now that $\mathfrak{G}$ is itself homomorphic to an isotope of $\mathfrak{G}'$ if there exists a mapping P of $\mathfrak{G}$ on $\mathfrak{G}'$ and nonsingular transformations S, T, U of $\mathfrak{G}$ such that $\mathfrak{G}'=\mathfrak{G}P$ and $(xP,yP)=(x\cdot y)P$ where $(x',y')=(x'S\cdot y'T)U$. This is equivalent to (17) with $A=PS$, $B=PT$, $C=PU^{-1}$. Then $A=C(US)$, $B=P(CT)$. Hence (17) implies that $\mathfrak{G}$ is homomorphic to an isotope of $\mathfrak{G}'$ if there exist nonsingular transformations S and T of $\mathfrak{G}'$ such that $A=CS$, $B=CT$. However such transformations may not exist.

For example let $\mathfrak{G}$ be any group of at least three distinct elements and C be any singular mapping of $\mathfrak{G}$ on a set $\mathfrak{G}'$ of at least two distinct elements. Then there exist distinct elements g, h, k in $\mathfrak{G}$ such that $gC=hC$, $gC\neq kC$. The transformation N on $\mathfrak{G}$ which interchanges h and k and leaves all other elements of $\mathfrak{G}$ unaltered is nonsingular and we define a homotopy of $\mathfrak{G}$ on $\mathfrak{G}'$

with $A=B=NC$. This defines the product operation in $\mathfrak{G}'$ and is such that $gN=g$, $gNC=gC\neq kN=hNC$. However if $A=CS$ we could conclude from $gC=hC$ that $gCS=hCS$, $gA=hA$, a contradiction.

Let us observe that if we designate the right and left multiplications of $\mathfrak{G}'$ respectively by $R_{xA}^{(1)}$ and $L_{yB}^{(1)}$ then (17) is equivalent to $xAR_{yB}^{(1)}=xR_yC$, $yBL_{xA}^{(1)}=yL_xC$ and thus to

$$AR_{yB}^{(1)} = R_yC, \qquad BL_{xA}^{(1)} = L_xC. \tag{18}$$

We assume that $\mathfrak{G}$ is a loop homotopic to a loop $\mathfrak{G}'$ and let e be the identity element of $\mathfrak{G}$. Replace x and y in (18) by e to obtain

$$C = AR_{eB}^{(1)} = BL_{eA}^{(1)}.$$

Since $\mathfrak{G}'$ is a loop both $R_{eB}^{(1)}$ and $L_{eA}^{(1)}$ are nonsingular transformations on $\mathfrak{G}'$ and $\mathfrak{G}$ is homomorphic to the isotope $\mathfrak{G}''$ of $\mathfrak{G}'$ defined by $(x', y')=x'S\cdot y'T$ where $S^{-1}=R_{eB}^{(1)}$, $T^{-1}=L_{eA}^{(1)}$. Since $(xC, yC)=(x\cdot y)C$ we see that eC is the identity element of $\mathfrak{G}''$ and $\mathfrak{G}''$ is a loop. We have proved

THEOREM 13. *Let a loop $\mathfrak{G}$ be homotopic to a loop $\mathfrak{G}'$. Then $\mathfrak{G}$ is homomorphic to a loop isotopic to $\mathfrak{G}'$.*

If $\mathfrak{H}$ is the normal divisor of $\mathfrak{G}$ which is mapped by C on eC the quotient loop $\mathfrak{G}/\mathfrak{H}$ is isotopic to $\mathfrak{G}'$. We note also that we have the almost immediate

COROLLARY. *Let a group $\mathfrak{G}$ be homotopic to a loop $\mathfrak{G}'$. Then $\mathfrak{G}$ is homomorphic to $\mathfrak{G}'$ and $\mathfrak{G}'$ is a group.*

For by Theorem 13 the group $\mathfrak{G}$ is homomorphic to $\mathfrak{G}''$, $\mathfrak{G}''$ is a group isotopic to $\mathfrak{G}'$, $\mathfrak{G}''$ is isomorphic to $\mathfrak{G}'$ by Theorem 2.

16. **Special loops.** A loop $\mathfrak{G}$ may be called *alternative* if the conditions

$$R_g^{-1} = R_{g^{-1}}, \qquad L_g^{-1} = L_{g^{-1}}, \qquad R_gL_gR_x = R_{g\cdot xL_g}, \qquad R_gL_{g\cdot x} = L_xR_gL_g$$

hold for every g and x of $\mathfrak{G}$ where $g\cdot g^{-1}=g^{-1}\cdot g=e$ is the identity element of $\mathfrak{G}$. Then it is known(5) that every two elements of $\mathfrak{G}$ generate a subloop which is a group.

An example of such a loop is given by the nonzero elements of the nonassociative Cayley-Dickson division algebra of order eight over the field of all real numbers. This loop contains a finite loop $\mathfrak{G}$ of sixteen elements. It consists of eight basal elements $e_1=1$, $e_2, \cdots, e_7$ and their negatives and $e_i(e_je_k)=\pm(e_ie_j)e_k$. The algebra is a central simple algebra and thus the center $\mathfrak{C}$ of the knot $\mathfrak{G}$ consists of the group of two elements 1, -1. Evidently the

(5) See R. Moufang, *Struktur von Alternativkorpern*, Math. Ann. vol. 110 (1935) pp. 416–430, for a discussion showing that this set of postulates imply that every two elements of the loop generate a group. The postulates are given also in the Hausman-Ore paper on page 993.

quotient loop $\mathfrak{G}/\mathfrak{C}$ is the abelian group of exponent two and order eight. This is important as an example of a case where $\mathfrak{H}$ is a group and $\mathfrak{G}/\mathfrak{H}$ is a group but $\mathfrak{G}$ is a loop which is not associative.

The theory of alternative loops is only one of many interesting questions about loops suggested by topics of the theory of nonassociative algebras. For example, the construction of new simple algebras recently(6) given by the author yields, by the use of basal units, a construction of what may be called a crossed extension $\mathfrak{B}$ of every finite loop $\mathfrak{A}$, in which $\mathfrak{A}$ is a normal divisor of $\mathfrak{B}$, and the quotient loop $\mathfrak{B}/\mathfrak{A}$ is a prescribed permutation group on $\mathfrak{A}$ leaving its identity element unaltered. The study of other properties of such loops should prove to be of considerable interest.

(6) *Non-associative algebras.* II. *New simple algebras*, Ann. of Math. vol. 43 (1942) pp. 708–723.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

PERMANENT CONFIGURATIONS IN THE n-BODY PROBLEM

BY

CARL HOLTOM

1. **Introduction.** The problem of two bodies for spheres, homogeneous in concentric layers and finite in size, was first solved in a geometrical way by Newton [14](1) about 1685. Euler [4] gave the first detailed analytical solution of the problem in 1744. In 1772 Lagrange [8] gave four particular solutions of the three-body problem in his prize memoir. All solutions of the two-body problem and Lagrange's particular solutions of the three-body problem belong to that special class called permanent configurations(2), which we we shall define presently for the case of n-bodies.

Consider n free particles in space which attract each other along lines joining them according to any function of the distance. If all the masses are projected with initial velocities, and thereafter are acted upon only by the force of attraction, their orbits will be space curves. The problem of determining the positions and velocities of the bodies at any later time is the problem of n bodies. *A configuration of n bodies is said to be permanent if, as the masses move in their respective orbits, the ratios of the mutual distances remain constant.* Such a configuration may change in size but not in shape.

The literature on permanent configuration problems contains two methods of approach, (1) that of establishing continuity between the configurations for $n-1$ and n-bodies, and (2) that of characterizing the configurations for any n by the necessary and sufficient conditions which the $n(n-1)/2$ mutual distances must satisfy. Collinear configurations were discovered for the case $n=3$ by Euler [5] and for any n by Lehmann-Filhès [9] and F. R. Moulton [13], the latter using the first approach. Noncollinear configurations for the case $n=3$ were discovered by Lagrange [8], who also treated the collinear case. Dziobek [3] discussed the general approach (2) and arrived at some results for the case $n=4$, which case was treated in detail by MacMillan and Bartky [12]. W. L. Williams [16], applying the same method, considered the noncollinear case $n=5$. Some noncollinear solutions of the n-body problem have been found by Hoppe [7], Andoyer [1], Longley [10] and Emilia Breglia [2], but all have some element of symmetry.

Space permanent configurations are the scarcest and have the undesirable property that the motion of all bodies is either toward their common center

Presented to the Society, September 13, 1943; received by the editors March 8, 1943 and, in revised form, July 8, 1943.

(1) Numbers in brackets refer to the bibliography at the end of the paper.

(2) MacMillan [11, pp. 71–72, 75–78].

of gravity, in which case the configuration lasts for a finite time, or in the the opposite direction, in which case the configuration is expanding without limit(3). In fact it can be shown that all permanent configurations are either space configurations of the dilating without rotation type or they are planar(4). We shall treat only the plane configurations, and in particular only those rotating without dilation, for if the law of attraction is the Newtonian law and there exists a permanent configuration in which the n bodies revolve in concentric circles about the common center of gravity, the bodies may also move in similar conics with the common center of gravity as foci(5).

In most cases mentioned above it has not been necessary to restrict the masses or mass ratios, but in this paper we shall, upon applying approach (1), make certain restrictions which are sufficient, though they may not be necessary, to insure continuity.

2. **The equations for a permanent configuration.** The problem of n bodies belongs in the field of differential equations. If we require that the ratios of the mutual distances remain constant, the problem becomes one of permanent configurations and belongs in the theory of implicit functions. As pointed out in §1, we need require only that the motion be circular, that is, the distances themselves remain constant.

Let the origin be taken at the center of mass of the system, and let ω denote the constant angular velocity of the n masses about the origin. The equations for a permanent configuration are(6)

$$\begin{aligned} &\sum_{j=1}^{n}{}' \frac{x_i - x_j}{(r_{ij})^3} m_j - \lambda x_i = 0, \\ &\sum_{j=1}^{n}{}' \frac{y_i - y_j}{(r_{ij})^3} m_j - \lambda y_i = 0, \end{aligned} \qquad i = 1, 2, \cdots, n, \tag{2.1}$$

where $\lambda = \omega^2$, $(r_{ij})^2 = (x_i - x_j)^2 + (y_i - y_j)^2$ and the prime on the summation denotes the sum for all j except j equal i. It is a problem now in implicit function theory to solve equations (2.1) for the $2n$ coordinates x_i and y_i in terms of the $n+1$ parameters m_i and λ.

3. **Existence of a point of libration.** The first step in establishing continuity between the solutions for $n-1$ and n bodies consists of proving the existence of at least one point of libration for the case of $n-1$ bodies, and this section is devoted to that purpose.

Suppose that for a set of positive values $m_1, m_2, \cdots, m_{n-1}, \lambda$, and for $n \geqq 4$, the first $n-1$ equations of each set (2.1) have a solution (x_i^0, y_i^0)

(3) The space configuration for $n=4$ was first noted by Lehmann-Filhès [9]. Cf. also MacMillan [11, p. 74], and Wintner [17, p. 279].

(4) Wintner [17, pp. 287 ff.].

(5) Wintner [17, p. 300], or MacMillan [11, p. 74].

(6) Wintner [17, p. 302].

$(i=1, 2, \cdots, n-1)$ and that these points $P_i^0=(x_i^0, y_i^0)$ are not collinear. If an infinitesimal mass m_n be added to the system at the point (x_n, y_n) then x_n and y_n must satisfy the equations

$$\begin{aligned}\phi_1(x_n, y_n) &= \sum_{j=1}^{n-1} \frac{(x_n - x_j^0)m_j}{[(x_n - x_j^0)^2 + (y_n - y_j^0)^2]^{3/2}} - \lambda x_n = 0,\\ \phi_2(x_n, y_n) &= \sum_{j=1}^{n-1} \frac{(y_n - y_j^0)m_j}{[(x_n - x_j^0)^2 + (y_n - y_j^0)^2]^{3/2}} - \lambda y_n = 0.\end{aligned} \tag{3.1}$$

The subscript n may be omitted, and the two equations

$$\phi_1(x, y) = 0, \qquad \phi_2(x, y) = 0, \tag{3.2}$$

are the equations of two algebraic plane curves which we shall call C_1 and C_2 respectively. Since the $n-1$ points P_i^0 satisfy both equations (3.1), both curves pass through each of the P_i^0. It will be shown that C_1 and C_2 have at least one real intersection other than the $n-1$ points P_i^0.

In order to determine the behavior of the curves in the neighborhood of any one of the points P_i^0, we may write the equations for C_1 and C_2 in the form

$$\begin{aligned}G_1(x, y) &= (x - x_i^0)m_i + [(x - x_i^0)^2 + (y - y_i^0)^2]^{3/2} f_1(x, y) = 0,\\ G_2(x, y) &= (y - y_i^0)m_i + [(x - x_i^0)^2 + (y - y_i^0)^2]^{3/2} f_2(x, y) = 0,\end{aligned}$$

where

$$\begin{aligned}f_1(x, y) &= \sum_{j=1}^{n-1} \frac{(x - x_j^0)m_j}{[(x - x_j^0)^2 + (y - y_j^0)^2]^{3/2}} - \lambda x,\\ f_2(x, y) &= \sum_{j=1}^{n-1} \frac{(y - y_j^0)m_j}{[(x - x_j^0)^2 + (y - y_j^0)^2]^{3/2}} - \lambda y.\end{aligned} \tag{3.3}$$

Since $f_1(x_i^0, y_i^0)=0$, and $f_2(x_i^0, y_i^0)=0$ from the ith equation in each set (2.1), it follows that $f_1(x, y)$ and $f_2(x, y)$ may be expanded in powers of $x-x_i^0$ and $y-y_i^0$, vanishing for $x=x_i^0$, $y=y_i^0$. Thus

$$\begin{aligned}\phi_1(x, y) &= \frac{(x - x_i^0)m_i}{[(x - x_i^0)^2 + (y - y_i^0)^2]^{3/2}}\\ &\quad + A_1(x - x_i^0) + B_1(y - y_i^0) + \cdots,\\ \phi_2(x, y) &= \frac{(y - y_i^0)m_i}{[(x - x_i^0)^2 + (y - y_i^0)^2]^{3/2}}\\ &\quad + A_2(x - x_i^0) + B_2(y - y_i^0) + \cdots.\end{aligned} \tag{3.4}$$

The slope of C_1 is

$$-\frac{\partial G_1/\partial x}{\partial G_1/\partial y}=\frac{m_i x+3[(x-x_i^0)^2+(y-y_i^0)^2]^{1/2}(x-x_i^0)f_1(x,y)+[(x-x_i^0)^2+(y-y_i^0)^2]^{3/2}[A_1+\cdots]}{3[(x-x_i^0)^2+(y-y_i^0)^2]^{1/2}(y-y_i^0)f_1(x,y)+[(x-x_i^0)^2+(y-y_i^0)^2]^{3/2}[B_1+\cdots]},$$

and the slope of C_2 is

$$-\frac{\partial G_2/\partial x}{\partial G_2/\partial y}=\frac{3[(x-x_i^0)^2+(y-y_i^0)^2]^{1/2}(x-x_i^0)f_2(x,y)+[(x-x_i^0)^2+(y-y_i^0)^2]^{3/2}[B_2+\cdots]}{m_i y+3[(x-x_i^0)^2+(y-y_i^0)^2]^{1/2}(y-y_i^0)f_2(x,y+[(x-x_i^0)^2+(y-y_i^0)^2]^{3/2}[A_2+\cdots]}.$$

On evaluating these slopes at (x_i^0, y_i^0) we find that C_1 passes through P_i^0 tangent to $x=x_i^0$ and C_2 passes through P_i^0 tangent to $y=y_i^0$.

We shall determine next on which side of the tangent line the curve lies in the neighborhood of P_i^0. Let $x-x_i^0=r\cos\theta$, $y-y_i^0=r\sin\theta$ then equations (3.2) and (3.4) become, in polar coordinates,

$$\psi_1(r,\theta) = (m_i/r^2 + A_1 r)\cos\theta + B_1 r\sin\theta + \cdots = 0,$$
$$\psi_2(r,\theta) = (m_i/r^2 + A_2 r)\cos\theta + B_2 r\sin\theta + \cdots = 0.$$

From (3.3) we note that $\partial f_1/\partial y \equiv \partial f_2/\partial x$, and hence $B_1=A_2$. Since m_i and r are positive we may choose r so small that both (m_i/r^3+A_1) and (m_i/r^3+B_2) are positive. Then as θ varies from 0 to 2π, $\psi_1(r,\theta)$ changes sign at $\theta=\arctan[(m_i/r^3+A_1)/(-B_1)]$, and ψ_2 changes sign at

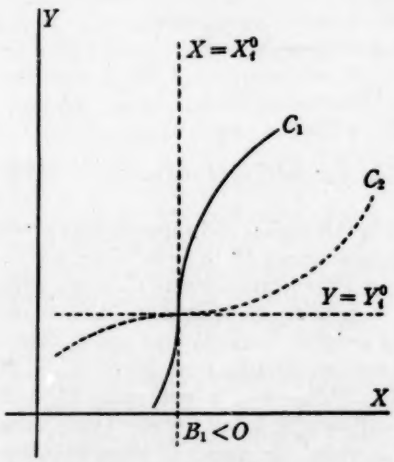


FIG. 1a

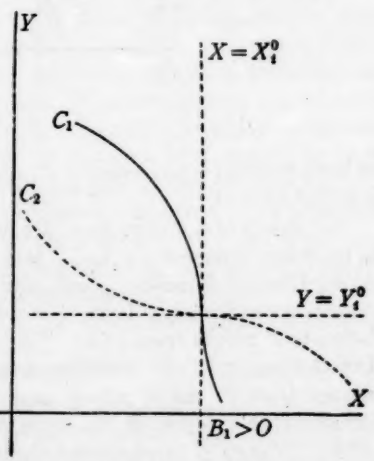


FIG. 1b

$\theta=\arctan[-B_1/(m_i/r^3+A_2)]$. Hence the curvature of both C_1 and C_2 at P_i^0 depends only upon the sign of B_1. The behavior of the curves at any P_i^0 is either that of Fig. 1a or Fig. 1b. In case $B_1=0$ the behavior of the curves

is governed by the sign of the first nonzero quantity in the sequence

$$\partial^2 f_1/\partial y^2 \equiv \partial^2 f_2/\partial x \partial y,\ \partial^2 f_1/\partial y^2 \equiv \partial^3 f_2/\partial y^2 \partial x, \cdots,$$

and the cases pictured in Figures 1a and 1b are all that can occur. In order to prove that there is at least one real solution of (3.2) other than the P_i^0, let us denote by P^0 that P_i^0 which has the largest ordinate. In case B_1 is negative at P^0 the curves are as shown in Fig. 2, and we shall prove that there is

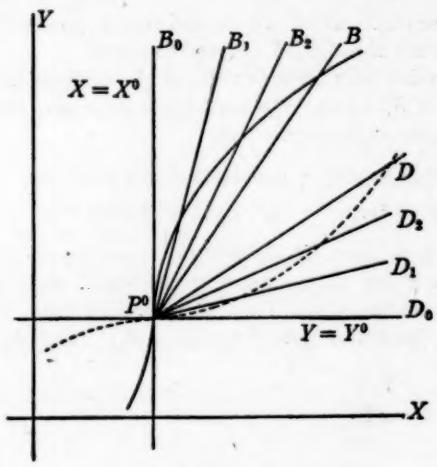


FIG. 2

at least one real intersection of C_1 and C_2 in that section of the plane for which $y > y^0$ and $x > x^0$.

By choice of notation there are no P_i^0 in this region and suppose that there is no other intersection. Let $P = (x, y)$ move from P^0 to $+\infty$ along $y = y^0$. Along this ray P meets C_2 first, namely at P^0, and then meets C_1 later, since ϕ_1 varies continuously from $+\infty$ to $-\infty$ as is easily verified in (3.1). Similarly, as P moves from P^0 to $+\infty$ along $x = x^0$ it meets C_1 first and C_2 later. Let this region of the plane be covered by rays leading from P^0 to ∞. As P moves from P^0 to ∞ along rays P^0D_1, P^0D_2, $\cdots$, it will meet C_2 first provided the slopes of the rays are sufficiently small, while along rays P^0B_1, P^0B_2, $\cdots$, it will meet C_1 first provided the slopes of these rays are sufficiently large. If there is no intersection of C_1 and C_2 in the region there is a last ray P^0D of the rays P^0D_i along which P meets C_2 first and a last ray P^0B of rays P^0B_i along which P meets C_1 first. The areas covered by the rays P^0D_i and P^0B_i may (a) overlap, (b) be adjacent, in which case P^0D and P^0B coincide, (c) be separated by a sector BP^0D.

Before taking up these three cases let us show that all rays from P^0 to ∞ that lie in the region $x>x^0$ and $y>y^0$ cut both curves at least once. From (3.1) it is seen that ϕ_1 and ϕ_2 are continuous functions of x and y for P within this region, and that ϕ_1 and ϕ_2 each vary continuously from $+\infty$ to $-\infty$ as P moves from P^0 to ∞ along any ray from P that lies in the region. Applying this result (a) to rays in the sector common to B_0P^0B and D_0P^0D, we conclude that every ray drawn from P^0 and lying in this sector, since it meets both curves and meets each curve first, must pass through an intersection of C_1 and C_2; (b) to the single ray common to both sectors, we reach the same conclusion as in (a); (c) to rays in the sector BP^0D separating those already covered, we find that every ray, since it does not meet C_1 first and does not meet C_2 first but does meet them both, must pass through an intersection of C_1 and C_2. In all cases one is led to a contradiction of the hypothesis that there is no intersection.

It is evident that the same proof can be applied in case B_1 is positive at P^0. In case there are several such points P^0 with the same ordinate the proof can be applied at any one of them yielding an intersection with ordinate greater than y^0. Thus we have the following theorem.

THEOREM 3.1. *There exists at least one point of libration at which a zero mass m_n can be placed, so that it will, together with the given $n-1$ positive masses, form a permanent configuration.*

4. **The solution as m_n becomes positive.** As shown in the last section there is a solution of (2.1) for positive λ, $m_1, m_2, \cdots, m_{n-1}$, and $m_n=0$. Let this solution be P_i^0. Since equations (2.1) are algebraic equations with coefficients functions of m_j, the solution functions x_i^0 and y_i^0 are continuous functions of m_n as long as the roots are finite and the equations do not have indeterminate forms. Consequently, x_i^0 and y_i^0 are continuous functions of m_n if no x_i^0 or no y_i^0 become infinite, or if no pair (x_i^0, y_i^0) becomes equal to another pair (x_k^0, y_k^0). Furthermore, the real roots x_i^0 and y_i^0 of algebraic equations (2.1) with real coefficients λ and m_i can disappear only by passing to infinity, or by an even number of real roots becoming complex conjugate quantities in pairs. Therefore, the problem is to determine as m_n varies whether

(1°) for all finite x_i^0 and y_i^0 any pair (x_j^0, y_j^0) can become equal to another pair (x_k^0, y_k^0),

(2°) any x_i^0 or y_i^0 can become infinite,

(3°) the solution (x_i^0, y_i^0) can ever become imaginary.

5. **The P_i^0 remain distinct.** This section is devoted to problem (1°) of the last section. We shall treat the more general case in which any m_j may vary.

Let the notation be chosen so that

$$x_1^0 \leqq x_2^0 \leqq x_3^0 \leqq \cdots \leqq x_{n-1}^0 \leqq x_n^0, \tag{5.1}$$

where there is at least one inequality in the noncollinear case. Suppose that

as m_j varies all P_i^0 remain finite, and P_j^0 approaches P_k^0 in a manner (a) such that $x_j^0 \neq x_k^0$, and that $j<k$. In the jth equation (2.1) there is a term involving $(x_j^0-x_k^0)/(r_{jk}^0)^3=\alpha_{jk}^0$ which becomes negatively infinite. If this jth equation is satisfied another term $\alpha_{jl}^0 m_l$ must become positively infinite, and according to our notation $l<j$. Now α_{jl}^0 appears besides only in the lth equation and becomes negatively infinite in the term $\alpha_{lj}^0 m_j$. If the lth equation is satisfied a term $\alpha_{lp}^0 m_p$ must become positively infinite and according to the notation $p<l$. Continuing in this manner one arrives eventually at the situation in which a term involving α_{1u}^0 with one of the subscripts 1 becomes negatively infinite. But all terms in the first equation are negative except $-\lambda x_1^0$, which cannot become positively infinite under the hypothesis that the P_i^0 remain finite. Hence, the hypothesis that P_j^0 approaches P_k^0 in the manner (a) is false.

Suppose that P_j^0 approaches P_k^0 in the manner (b) such that $x_j^0=x_k^0$. For this case let us choose the notation so that

$$y_1^0 \leqq y_2^0 \leqq y_3^0 \leqq \cdots \leqq y_{n-1}^0 \leqq y_n^0, \tag{5.2}$$

where again there is at least one inequality. In this notation, if P_j^0 becomes P_p^0 and P_k^0 becomes P_q^0, our hypothesis is that P_p^0 approaches P_q^0 in the manner $x_p^0=x_q^0$ and $p<q$. There is a term in the $(n+p)$th equation (2.1) involving $(y_p^0-y_q^0)/(r_{pq}^0)^3=\beta_{pq}^0$, which becomes negatively infinite. By an argument similar to that of the last paragraph one eventually arrives at the situation in which a term involving β_{1u}^0 with one of the subscripts 1 becomes negatively infinite. But all terms in the $(n+1)$th equation are negative except $-\lambda y_1^0$, which cannot become positively infinite under the hypothesis that the P_i^0 remain finite. Hence, the hypothesis that P_j^0 approaches P_k^0 in the manner (b) is false. This completes the proof of the following theorem.

THEOREM 5.1. *As the m_i vary and the solution functions x_i^0 and y_i^0 remain finite the P_i^0 remain distinct.*

6. **The P_i^0 remain finite.** In order to show that as any m_j varies no x_i^0 or y_i^0 can become infinite, let us again adopt the notation (5.1) and suppose that x_j^0 becomes positively infinite. From the center of gravity equation

$$\sum_{i=1}^{n} m_i x_i^0 = 0, \tag{6.1}$$

some x_k^0 must become negatively infinite, and from (5.1) it follows that x_1^0 approaches $-\infty$ and x_n^0 approaches $+\infty$. Consider the first equation (2.1). In order that it be satisfied P_2^0 must approach P_1^0 in a manner such that $x_1^0 \neq x_2^0$, that is, x_2^0 must approach $-\infty$. Now consider the second equation. If it is to be satisfied, P_3^0 must approach P_2^0 in a manner such that $x_2^0 \neq x_3^0$, that is, x_3^0 must approach $-\infty$. Continuing this process we are led to the conclusion that all x_i^0 approach $-\infty$, hence the hypothesis must be false.

In order to show that no y_i^0 can become infinite, we adopt the notation (5.2) and suppose that y_i^0 becomes positively infinite as m_j varies. From the center of gravity equation

$$\sum_{i=1}^{n} m_i y_i^0 = 0, \tag{6.2}$$

and the notation (5.2), it follows that y_1^0 approaches $-\infty$, and y_n^0 approaches $+\infty$. Consider the $(n+1)$th, $(n+2)$th, and so on of equations (2.1) and the proof is the same as that in the preceding paragraph with the corresponding changes in notation. Thus we have the following theorem.

THEOREM 6.1. *As the m_i vary the P_i^0 remain finite.*

7. **The solution as m_n vanishes.** Before we take up the question of the solution becoming imaginary (3° of §4), it is to be shown that as m_n approaches zero the equations (2.1) and the solution functions x_i^0 and y_i^0 remain regular, and that there is, accordingly, only one limiting position of the P_i^0 for the value $m_n=0$. Let a solution for all positive m_i and λ be P_i', and let some m_j approach zero.

First, if some P_j' approaches some P_k' in any manner then it must be that P_j' approaches P_{j-1}' or P_{j+1}', for all other possibilities lead to the fact that one of the equations cannot be satisfied by arguments used in §§5 and 6.

Secondly, if P_j' approaches P_{j+1}' in the manner $x_j' = x_{j+1}'$, and if the notation is that of (5.2) the $(n+j)$th equation (2.1) cannot be satisfied unless P_{j-1}' approaches P_j' in the manner $x_{j-1}' = x_j'$. Thus P_{j-1}' approaches P_{j+1}' in the manner $x_{j-1}' = x_{j+1}'$ as m_j vanishes. But then the $(n+j+1)$th and $(n+j-1)$th equations cannot be satisfied unless P_{j-2}' approaches P_{j-1}' in the manner $x_{j-2}' = x_{j-1}'$ and P_{j+2}' approaches P_{j+1}' in the manner $x_{j+2}' = x_{j+1}'$. This shifts the difficulty to the $(n+j+2)$th and $(n+j-2)$th equations. On continuing this process one arrives at the $(n+1)$th and $(2n)$th equations which cannot be satisfied under the hypothesis it was necessary to make.

In case P_j' approaches P_{j+1}' in the manner $x_j' \neq x_{j+1}'$, then, if the notation is that of (5.1), the jth equation (2.1) cannot be satisfied unless P_{j-1}' approaches P_j' in the manner $x_{j-1}' \neq x_j'$ as m_j vanishes. Thus P_{j-1}' approaches P_{j+1}' in the manner $x_{j-1}' \neq x_{j+1}'$. But the $(j+1)$th and $(j-1)$th equations cannot be satisfied unless P_{j-2}' approaches P_{j-1}' in the manner $x_{j-2}' \neq x_{j-1}'$, and P_{j+2}' approaches P_{j+1}' in the manner $x_{j+2}' \neq x_{j+1}'$. This shifts the difficulty to the $(j-2)$th and $(j+2)$th equations, and on continuing this process one eventually arrives at the 1st and nth equations which cannot be satisfied under the hypothesis it was necessary to make.

In the third place, if any x_i' or y_i' becomes positively infinite as m_j vanishes, it must be that x_j' or y_j' becomes positively infinite, for all other possibilities lead to the fact that all x_i' or all y_i' become negatively infinite

by arguments used in §6. If the notation is that of (5.1), the jth equation cannot be satisfied as x_j' becomes positively infinite unless P_j' approaches P_{j-1}' in the manner $x_j' \neq x_{j-1}'$, that is, x_{j-1}' must approach $+\infty$. Then the $(j-1)$th equation cannot be satisfied unless P_{j-1}' approaches P_{j-2}' in the manner $x_{j-1}' \neq x_{j-2}'$, in which case x_{j-2}' approaches $+\infty$. Continuing this process one arrives at the conclusion that $x_1', x_2', \cdots, x_{j-1}', x_j'$ become positively infinite, and by choice of notation $x_{j+1}', x_{j+2}', \cdots, x_n'$ become positively infinite. Hence the center of gravity equation (6.1) cannot be satisfied, and the hypothesis that x_j' become positively infinite is false.

The proof that y_j' cannot become positively infinite as m_j vanishes is precisely the same as that of the last paragraph with the notation that of (5.2) and the argument beginning with the consideration of the $(n+j)$th equation (2.1). The proof that no x_j' or y_j' can become negatively infinite as m_j vanishes will read the same as the proof above with the words "positively infinite" replaced by "negatively infinite." This concludes the proof of the following:

THEOREM 7.1. *Any solution P_i' of* (2.1) *for positive m_i and λ remains regular as any m_i vanishes.*

In particular we shall use the following corollary.

COROLLARY 7.1. *The solution P_i' for positive m_i and λ remains regular as m_n vanishes.*

8. **The condition that the P_i^0 remain real.** Since the solutions of (2.1) are continuous functions of m_n, it follows that no two solutions which are real for $m_n = 0$ can ever become complex conjugate solutions for any positive value of m_n without having first become equal. If a multiple solution of (2.1) is impossible for a set of finite positive values of m_i, then it is impossible for any real solution to disappear by becoming complex or for any complex solution to become real.

The conditions that a set of simultaneous algebraic equations shall have a multiple solution are that a set of values of the variable shall satisfy the equations, and that the Jacobian of the functions with respect to the independent variables shall vanish for the same set of values.

Since equations (2.1) are invariant under a rotation of axes, let the axes be chosen so that $y_2 = 0$ in case $n = 2$, and so that $y_1 = 0$ for $n \geqq 3$. It is to be shown that for $n = 2$ the first three equations (2.1) are independent for all positive values of λ, m_1 and m_2, while for $n \geqq 3$ the $2n-1$ equations obtained from (2.1) by omitting the $(n+1)$th form an independent set for any positive λ, m_1, m_2, and sufficiently small positive $m_3, m_4, \cdots, m_n$.

Let F_i and F_{n+i} denote the left members of (2.1) and let the Jacobian of these functions be Δ_n. Furthermore, let the minor obtained by deleting the $(n+1)$th row and $(n+1)$th column of Δ_n be D_n and let $n \geqq 3$. A real solution P_i^0 of (2.1) for positive $\lambda, m_1, m_2, \cdots, m_{n-1}$, and $m_n = 0$ will remain real as m_n

increases from zero through positive values for which D_n does not vanish. This will be established when it is shown that $D_n \neq 0$ for sufficiently small positive m_j, $j=3, 4, \cdots, n$.

The following notation will be used. $\partial F_i/\partial x_j = a_{ij}$, $\partial F_{n+i}/\partial y_j = b_{ij}$, $\partial F_i/\partial y_j = \partial F_{n+i}/\partial x_j = c_{ij}$. From (2.1) it follows that, for $i \neq j$,

$$a_{ij} = \frac{3(x_i - x_j)^2 - (r_{ij})^2}{(r_{ij})^5} m_j, \qquad b_{ij} = \frac{3(y_i - y_j)^2 - (r_{ij})^2}{(r_{ij})^5} m_j,$$

$$c_{ij} = \frac{3(x_i - x_j)(y_i - y_j)}{(r_{ij})^5} m_j, \tag{8.1}$$

and, for $i=j$,

$$a_{ii} = -\sum_{k=1}^{n-1} a_{i\,i+k} - \lambda, \quad b_{ii} = -\sum_{k=1}^{n-1} b_{i\,i+k} - \lambda, \quad c_{ii} = -\sum_{k=1}^{n-1} c_{i\,i+k}, \tag{8.2}$$

where $a_{il+n} = a_{il}$, $b_{il+n} = b_{il}$, $c_{il+n} = c_{il}$.

9. **Case $n=2$.** Equations (2.1) in this case consist of four equations. Let us choose the axes so that $y_2 = 0$, then the solution of these equations is

$$\begin{aligned} x_1 &= -m_2/(\lambda M^2)^{1/3}, & y_1 &= 0, \\ x_2 &= m_1/(\lambda M^2)^{1/3}, & y_2 &= 0, \end{aligned} \tag{9.1}$$

where $M = m_1 + m_2$.

To compute the elements of Δ_2 the following quantities, $x_1 - x_2 = -(M/\lambda)^{1/3}$, $y_1 - y_2 = 0$, $r_{12} = (M/\lambda)^{1/3}$, are substituted in (8.1) and (8.2). We find

$$\Delta_2 = \begin{vmatrix} -2\lambda m_2/M - \lambda & 2\lambda m_2/M & 0 & 0 \\ 2\lambda m_1/M & -2\lambda m_1/M - \lambda & 0 & 0 \\ 0 & 0 & -\lambda m_1/M & -\lambda m_2/M \\ 0 & 0 & -\lambda m_1/M & -\lambda m_2/M \end{vmatrix}.$$

It is quite evident that $\Delta_2 = 0$ and it remains to be shown that the minor D_2 in the upper left-hand corner is different from zero. After adding the second column to the first, and then subtracting the first row from the second,

$$D_2 = \begin{vmatrix} -\lambda & 2\lambda m_2/M & 0 \\ 0 & -3\lambda & 0 \\ 0 & 0 & -\lambda m_1/M \end{vmatrix}.$$

Since $D_2 \neq 0$ for all finite positive values of λ, m_1 and m_2 the solution of (2.1) for $m_2 = 0$, namely, $x_1 = 0$, $y_1 = 0$, $x_2 = (m_1/\lambda)^{1/3}$, $y_2 = 0$, remains real and varies continuously to the values given in (9.1) for all finite positive values of λ, m_1 and m_2.

10. **Case $n=3$.** By choosing the axes so that $y_1 = 0$ the solution of the six equations (2.1) is

$$
\begin{aligned}
x_1 &= 2(m_2^2 + m_2m_3 + m_3^2)K, & y_1 &= 0,\\
x_2 &= -(m_2m_3 + m_1m_3 + 2m_1m_2 - m_3^2)K, & y_2 &= -3^{1/2}M\,m_3K,\\
x_3 &= -(2m_1m_3 + m_1m_2 + m_2m_3 - m_2^2)K, & y_3 &= 3^{1/2}M\,m_2K,
\end{aligned} \tag{10.1}
$$

where $M=m_1+m_2+m_3$, and $K=1/(2\lambda^{1/3}M^{2/3}(m_2^2+m_2m_3+m_3^2)^{1/2})$.

If the elements of Δ_3 are computed in terms of λ, m_1, m_2 and m_3, Δ_3 will vanish, as can be shown by direct computation, since the system (2.1) is dependent and y_1 can be chosen arbitrarily. D_3 is the minor obtained by leaving out the 4th row and 4th column of Δ_3 and has determinant different from zero for all finite positive λ, m_1, m_2, and m_3, as a rather long series of computations will show.

However, the proof that D_3 is different from zero for all λ, m_1, m_2 and sufficiently small m_3 is short and will be given here, since it is the same type of proof that will be used in the general case. In the notation (8.1) and (8.2),

$$
\Delta_3 = \begin{vmatrix}
a_{11} & a_{12} & a_{13} & c_{11} & c_{12} & c_{13}\\
a_{21} & a_{22} & a_{23} & c_{21} & c_{22} & c_{23}\\
a_{31} & a_{32} & a_{33} & c_{31} & c_{32} & c_{33}\\
c_{11} & c_{12} & c_{13} & b_{11} & b_{12} & b_{13}\\
c_{21} & c_{22} & c_{23} & b_{21} & b_{22} & b_{23}\\
c_{31} & c_{32} & c_{33} & b_{31} & b_{32} & b_{33}
\end{vmatrix}, \quad
D_3 = \begin{vmatrix}
a_{11} & a_{12} & a_{13} & c_{12} & c_{13}\\
a_{21} & a_{22} & a_{23} & c_{22} & c_{23}\\
a_{31} & a_{32} & a_{33} & c_{32} & c_{33}\\
c_{21} & c_{22} & c_{23} & b_{22} & b_{23}\\
c_{31} & c_{32} & c_{33} & b_{32} & b_{33}
\end{vmatrix}. \tag{10.2}
$$

After rearranging rows and columns in Δ_3 we have

$$
\Delta_3 = \begin{vmatrix}
a_{11} & a_{12} & c_{11} & c_{12} & a_{13} & c_{13}\\
a_{21} & a_{22} & c_{21} & c_{22} & a_{23} & c_{23}\\
c_{11} & c_{12} & b_{11} & b_{12} & c_{13} & b_{13}\\
c_{21} & c_{22} & b_{21} & b_{22} & c_{23} & b_{23}\\
a_{31} & a_{32} & c_{31} & c_{32} & a_{33} & c_{33}\\
c_{31} & c_{32} & b_{31} & b_{32} & c_{33} & b_{33}
\end{vmatrix},
$$

and if $m_3=0$, Δ_3 and D_3 become

$$
\Delta_3 = \left|\begin{array}{cccc:cc}
 & & & & 0 & 0\\
 & \Delta_2 & & & 0 & 0\\
 & & & & 0 & 0\\
 & & & & 0 & 0\\
\hdashline
a_{31} & a_{32} & c_{31} & c_{32} & a_{33} & c_{33}\\
c_{31} & c_{32} & b_{31} & b_{32} & c_{33} & b_{33}
\end{array}\right|,
$$

and

$$(10.3)\quad D_3=\begin{vmatrix} -\left(\frac{2m_2}{M}+1\right)\lambda & \frac{2m_2}{M}\lambda & 0 & 0 & 0 \\ \frac{2m_1}{M}\lambda & -\left(\frac{2m_1}{M}+1\right)\lambda & 0 & 0 & 0 \\ -\frac{m_1}{4M}\lambda & -\frac{m_2}{M}\lambda & -\frac{3}{4}\lambda & \frac{-3^{1/2}m_2}{M}\lambda & \frac{-3^{1/2}(m_1-m_2)}{M}\lambda \\ 0 & 0 & 0 & -\frac{m_2}{M}\lambda & 0 \\ \frac{3^{1/2}m_1}{M}\lambda & \frac{-3^{1/2}m_2}{M}\lambda & \frac{-3^{1/2}(m_1-m_2)}{M}\lambda & \frac{5m_2}{4M}\lambda & -\frac{9}{4}\lambda \end{vmatrix}.$$

On expanding (10.3) we obtain

$$(10.4)\qquad D_3=-(3^4m_1m_2^2/4M^5)\lambda^5.$$

Since D_3 of (10.2) is a continuous function of m_3 and, as (10.4) reveals, is different from zero for positive λ, m_1, m_2, and $m_3=0$, it is different from zero for sufficiently small positive m_3.

11. **Case $n=4$.** Using the notation (8.1) and (8.2) we have

$$\Delta_4=\left|\begin{array}{c|c} a_{ij} & c_{ij} \\ \hline c_{ij} & b_{ij} \end{array}\right|, \qquad i,j=1,2,3,4,$$

and

$$D_4=\left|\begin{array}{cccc|ccc} & & & & c_{12} & c_{13} & c_{14} \\ & & a_{ij} & & c_{22} & c_{23} & c_{24} \\ & & & & c_{32} & c_{33} & c_{34} \\ & & & & c_{42} & c_{43} & c_{44} \\ \hline c_{21} & c_{22} & c_{23} & c_{24} & b_{22} & b_{23} & b_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} & b_{32} & b_{33} & b_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} & b_{42} & b_{43} & b_{44} \end{array}\right|, \qquad i,j=1,2,3,4.$$

After rearranging rows and columns, we have

$$\Delta_4=\left|\begin{array}{cccccc|cc} & & & & & & a_{14} & c_{14} \\ & & & & & & a_{24} & c_{24} \\ & & & \Delta_3 & & & a_{34} & c_{34} \\ & & & & & & c_{14} & b_{14} \\ & & & & & & c_{24} & b_{24} \\ & & & & & & c_{34} & b_{34} \\ \hline a_{41} & a_{42} & a_{43} & c_{41} & c_{42} & c_{43} & a_{44} & c_{44} \\ c_{41} & c_{42} & c_{43} & b_{41} & b_{42} & b_{43} & c_{44} & b_{44} \end{array}\right|, \qquad D_4=\left|\begin{array}{cccccc|cc} & & & & & & a_{14} & c_{14} \\ & & & & & & a_{24} & c_{24} \\ & & & D_3 & & & a_{34} & c_{34} \\ & & & & & & c_{24} & b_{24} \\ & & & & & & c_{34} & b_{34} \\ \hline a_{41} & a_{42} & a_{43} & c_{42} & c_{43} & & a_{44} & c_{44} \\ c_{41} & c_{42} & c_{43} & b_{42} & b_{43} & & c_{44} & b_{44} \end{array}\right|.$$

For $m_4=0$ these become

$$\Delta_4 = \left|\begin{array}{c:c} \Delta_3 & \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} \\ \hline \begin{matrix} a_{41} & a_{42} & a_{43} & c_{41} & c_{42} & c_{43} \\ c_{41} & c_{42} & c_{43} & b_{41} & b_{42} & b_{43} \end{matrix} & \begin{matrix} a_{44} & c_{44} \\ c_{44} & b_{44} \end{matrix} \end{array}\right|, \quad D_4 = \left|\begin{array}{c:c} D_3 & \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} \\ \hline \begin{matrix} a_{41} & a_{42} & a_{43} & c_{42} & c_{43} \\ c_{41} & c_{42} & c_{43} & b_{42} & b_{43} \end{matrix} & \begin{matrix} a_{44} & c_{44} \\ c_{44} & b_{44} \end{matrix} \end{array}\right|.$$

Since x_i and y_i are continuous functions of m_4, D_4 is a continuous function of m_4 and is different from zero for $m_4=0$, provided that λ, m_1 and m_2 are positive, that m_3 is so chosen that $D_3 \neq 0$, and that (x_4, y_4) is not a multiple solution of the 4th and 8th equations (2.1). Hence, $D_4 \neq 0$ for sufficiently small positive values of m_4, and the solution, real for positive λ, m_1, m_2, m_3, and $m_4=0$, varies continuously and remains real for sufficiently small positive values of m_4.

12. **Existence of a solution for positive m_n.** Continuing the process of §§9, 10, 11 it is possible for one to choose $m_5, m_6, \cdots, m_{n-1}$ successively so that $D_5, D_6, \cdots, D_{n-1}$ remain different from zero. In the general case Δ_n will be zero due to the dependence of equations (2.1). The minor D_n obtained from Δ_n by deleting the $(n+1)$th row and $(n+1)$th column will be, in the notation of (8.1) and (8.2) and after rows and columns have been rearranged,

$$D_n = \left|\begin{array}{c:c} D_{n-1} & \begin{matrix} a_{1n} & c_{1n} \\ a_{2n} & c_{2n} \\ \cdot & \cdot \\ \cdot & \cdot \\ a_{n-1\,n} & c_{n-1\,n} \\ c_{2n} & b_{2n} \\ c_{3n} & b_{3n} \\ \cdot & \cdot \\ \cdot & \cdot \\ c_{n-1\,n} & b_{n-1\,n} \end{matrix} \\ \hline \begin{matrix} a_{n1} & a_{n2} \cdots a_{n\,n-1} & c_{n2} & c_{n3} \cdots c_{n\,n-1} \\ c_{n1} & c_{n2} \cdots c_{n\,n-1} & b_{n2} & b_{n3} \cdots b_{n\,n-1} \end{matrix} & \begin{matrix} a_{nn} & c_{nn} \\ c_{nn} & b_{nn} \end{matrix} \end{array}\right|.$$

For $m_n=0$ we have

$$D_n = D_{n-1} \left|\begin{matrix} \partial F_n/\partial x_n & \partial F_n/\partial y_n \\ \partial F_{2n}/\partial x_n & \partial F_{2n}/\partial y_n \end{matrix}\right|.$$

Since x_i and y_i are continuous functions of m_n, D_n is a continuous function

of m_n; and, if (x_n, y_n) is not a multiple solution of the nth and $(2n)$th equations (2.1), $D_n \neq 0$ for $m_n = 0$. Therefore, $D_n \neq 0$ for sufficiently small values of m_n, and the solution, real for positive λ, m_1, m_2, $\cdots$, m_{n-1}, and $m_n = 0$, varies continuously and remains real for sufficiently small positive values of m_n. This concludes the proof of the following theorem.

THEOREM 12.1. *For arbitrary finite positive λ, m_1, m_2, and sufficiently small positive m_3, m_4, $\cdots$, m_n there exists at least one real solution of equations* (2.1).

It should be noted that, from the Theorem of Lagrange and our remark in §10, m_3 need not be restricted and hence we have the slightly more general theorem.

THEOREM 12.2. *For arbitrary finite positive masses m_1, m_2, m_3, arbitrary finite angular velocity, and sufficiently small positive masses m_4, m_5, $\cdots$, m_n, there exists at least one noncollinear plane permanent configuration.*

13. **A special form for the equations.** In §12 we showed that under certain conditions $D_n \neq 0$. It follows from this result that for any λ, m_1, m_2, and properly chosen m_3, m_4, $\cdots$, m_n, equations (2.1) with the $(n+1)$th deleted form an independent set. It is easy to show, by processes similar to those of §§8 to 12, that the first and $(n+2)$th may be replaced by the center of gravity equations (6.1) and (6.2) yielding an independent set of $2n-1$ equations equivalent to the one just mentioned.

It is our purpose in this section to put this latter set of $2n-1$ equations into a form convenient for calculating a solution. We shall then solve them for the coordinates in terms of λ and the masses m_i. Let the mass ratios be $1, m_2/m_1 = \mu, m_{i+2}/m_1 = \sigma_i$ $(i = 1, 2, \cdots, n-2)$ and for the particular solution we wish to construct we shall take $\lambda = m_1$. The equations then become

$$
\begin{aligned}
&x_1 + \mu x_2 + \sum x_{2+j}\sigma_j = 0,\\
&y_1 + \mu y_2 + \sum y_{2+j}\sigma_j = 0,\\
&-x_2 + \frac{x_2 - x_1}{(r_{21})^3} + \sum \frac{x_2 - x_{2+j}}{(r_{2\,2+j})^3}\sigma_j = 0,\\
(13.1)\qquad &-x_i + \frac{x_i - x_1}{(r_{i1})^3} + \frac{x_i - x_2}{(r_{i2})^3}\mu + \sum{}^{*} \frac{x_i - x_{2+j}}{(r_{i\,2+j})^3}\sigma_j = 0,\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 3, 4, \cdots, n\\
&-y_i + \frac{y_i - y_1}{(r_{i1})^3} + \frac{y_i - y_2}{(r_{i2})^3}\mu + \sum{}^{*} \frac{y_i - y_{2+j}}{(r_{i\,2+j})^3}\sigma_j = 0,
\end{aligned}
$$

where $\sum$ denotes $\sum_{j=1}^{n-2}$ and $\sum^{*}$ denotes the sum over the same j except for $j = i-2$.

Let us multiply the next to last equation by y_i and the last by $-x_i$ and

form the sum. The resulting equation,

$$\frac{x_iy_1 - x_1y_i}{(r_{i1})^3} + \frac{x_iy_2 - x_2y_i}{(r_{i2})^3}\mu + \sum{}^* \frac{x_iy_{2+j} - x_{2+j}y_i}{(r_{i\,2+j})^3}\sigma_j = 0,$$

is used to replace the next to last above. By means of the complex numbers

$$z_{ij} = (y_i - y_j + (-1)^{1/2}(x_iy_j - x_jy_i))/(r_{ij})^3$$

we are now able to write the $2n-1$ equations (13.1) in the form

$$\begin{aligned}
&x_1 + \mu x_2 + \sum x_{2+j}\sigma_j = 0,\\
&y_1 + \mu y_2 + \sum y_{2+j}\sigma_j = 0,\\
&-x_2 + \frac{x_2 - x_1}{(r_{21})^3} + \sum \frac{x_2 - x_{2+j}}{(r_{2\,2+j})^3}\sigma_j = 0,\\
&-y_i + z_{i1} + z_{i2}\mu + \sum{}^* z_{i\,2+j}\sigma_j = 0, \qquad i = 3, 4, \cdots, n,
\end{aligned} \tag{13.2}$$

which is convenient for computing the solution.

In the case $n=4$ the coordinates may be expressed as power series in the mass ratios σ_1 and σ_2 with coefficients which are Laurent series in μ. For $n>4$ the coordinates may be expressed as power series in $(\sigma_i/3)^{1/3}$ with coefficients Laurent series in μ.

14. **Solution for $n=4$.** In this case the seven equations (13.2) are

$$\begin{aligned}
&x_1 + \mu x_2 + \sigma_1 x_3 + \sigma_2 x_4 = 0,\\
&y_1 + \mu y_2 + \sigma_1 y_3 + \sigma_2 y_4 = 0,\\
&-x_2 + \frac{x_2 - x_1}{(r_{21})^3} + \frac{x_2 - x_3}{(r_{23})^3}\sigma_1 + \frac{x_2 - x_4}{(r_{24})^3}\sigma_2 = 0,\\
&\frac{x_3y_1 - x_1y_3}{(r_{31})^3} + \frac{x_3y_2 - x_2y_3}{(r_{32})^3}\mu + \frac{x_3y_4 - x_4y_3}{(r_{34})^3}\sigma_2 = 0,\\
&-y_3 + \frac{y_3 - y_1}{(r_{31})^3} + \frac{y_3 - y_2}{(r_{32})^3}\mu + \frac{y_3 - y_4}{(r_{34})^3}\sigma_2 = 0,\\
&\frac{x_4y_1 - x_1y_4}{(r_{41})^3} + \frac{x_4y_2 - x_2y_4}{(r_{42})^3}\mu + \frac{x_4y_3 - x_3y_4}{(r_{43})^3}\sigma_1 = 0,\\
&-y_4 + \frac{y_4 - y_1}{(r_{41})^3} + \frac{y_4 - y_2}{(r_{42})^3}\mu + \frac{y_4 - y_3}{(r_{43})^3}\sigma_1 = 0.
\end{aligned} \tag{14.1}$$

Since one of the coordinates is arbitrary we may choose $y_1=0$. The solution of these equations may be obtained as follows. With $\sigma_2=0$ the first five equations (14.1) form an independent set of the six equations (2.1), and the solution functions may be obtained from (10.1) by proper change of parame-

ters. The solution functions thus obtained and arranged in powers of σ_1 are

$$
\begin{aligned}
x_1 &= \mu(1+\mu)^{-2/3} + \frac{3-\mu}{6}(1+\mu)^{-5/3}\sigma_1 + \cdots, \\
x_2 &= -(1+\mu)^{-2/3} + \frac{1-3\mu}{6}(1+\mu)^{-5/3}\sigma_1 + \cdots, \\
y_2 &= \frac{3^{1/2}}{2}\left[0 - \frac{(1+\mu)^{1/3}}{\mu}\sigma_1 + \cdots\right], \\
x_3 &= -\frac{1-\mu}{2}(1+\mu)^{-2/3} - \frac{9+14\mu+13\mu^2}{12\mu}(1+\mu)^{-5/3}\sigma_1 + \cdots, \\
y_3 &= \frac{3^{1/2}}{2}\left[(1+\mu)^{1/3} - \frac{3+\mu}{6\mu}(1+\mu)^{-2/3}\sigma_1 + \cdots\right].
\end{aligned} \tag{14.2}
$$

The last two equations (14.1) may be obtained from the preceding two by the transformation, in two rowed notation,

$$
\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & y_2 & y_3 & y_4 & \mu & \sigma_1 & \sigma_2 \\ x_1 & x_2 & x_4 & x_3 & -y_2 & -y_4 & -y_3 & \mu & \sigma_2 & \sigma_1 \end{pmatrix}, \tag{14.3}
$$

and then setting $\sigma_1 = 0$. Hence the solution functions x_4, y_4 of the last two equations may be obtained from the solution functions x_3, y_3 except for the coefficients of the powers of σ_1. On substituting

$$
\begin{aligned}
x_4 &= -((1-\mu)/2)(1+\mu)^{-2/3} + c_1(\mu)\sigma_1 + \cdots, \\
y_4 &= -(3^{1/2}/2)\left[(1+\mu)^{1/3} + c_2(\mu)\sigma_1 + \cdots\right]
\end{aligned}
$$

in the last two equations (14.1) and equating coefficients we find

$$
\begin{aligned}
c_1(\mu) &= \left(\frac{3^{1/2}}{27} + \frac{5}{12}\right)\frac{1}{\mu} - \left(\frac{2(3)^{1/2}}{81} - \frac{29}{36}\right) \\
&\quad - \left(\frac{3^{1/2}}{81} + \frac{7}{9}\right)\mu + \cdots, \\
c_2(\mu) &= \left(\frac{2(3)^{1/2}}{81} + \frac{5}{18}\right)\frac{1}{\mu} + \left(\frac{8(3)^{1/2}}{243} + \frac{11}{54}\right) \\
&\quad + \left(\frac{2(3)^{1/2}}{81} - \frac{8}{27}\right)\mu + \cdots.
\end{aligned} \tag{14.4}
$$

This completes the solution of (14.1) as functions of μ and σ_1.

In order to obtain the coefficients of σ_2 in the solution functions of (14.1) we may use the fact that the system (14.1) remains unchanged by the transformation (14.3) and that the solution functions must have this same prop-

erty. Coefficients of higher powers of the parameters may be computed, and herein lies the convenience of the form (14.1). The coefficients of any term $\sigma_1^p\sigma_2^q$ in the expansion of $x_1, y_2, x_2, x_3, y_3, x_4, y_4$ may be determined successively, and in this order, by substituting series with undetermined coefficients into the equations in the order given in (14.1).

The solution for $n=4$, computed as described above and arranged in powers of σ_2, will follow immediately. The coefficients in this series are themselves arranged in powers of σ_1 with coefficients which are Laurent series in μ.

$$
\begin{aligned}
x_1 &= \mu(1+\mu)^{-2/3} + \frac{3-\mu}{6}(1+\mu)^{-5/3}\sigma_1 + \cdots + \left\{\frac{3-\mu}{6}(1+\mu)^{-5/3}\right. \\
&\quad \left. + \left[-\left(\frac{2(3)^{1/2}}{27} + \frac{1}{12}\right)\frac{1}{\mu} + \cdots\right]\sigma_1 + \cdots\right\}\sigma_2 + \cdots, \\
x_2 &= -(1+\mu)^{-2/3} + \frac{1-3\mu}{6}(1+\mu)^{-5/3}\sigma_1 + \cdots + \left\{\frac{1-3\mu}{6}(1+\mu)^{-5/3}\right. \\
&\quad \left. + \left[-\frac{3}{4\mu^2} + \cdots\right]\sigma_1 + \cdots\right\}\sigma_2 + \cdots, \\
y_2 &= \frac{3^{1/2}}{2}\left(0 - \frac{(1+\mu)^{1/3}}{\mu}\sigma_1 + \cdots + \left\{\frac{(1+\mu)^{1/3}}{\mu}\right.\right. \\
&\quad \left.\left. + \left[\frac{0}{\mu^2} + \cdots\right]\sigma_1 + \cdots\right\}\sigma_2 + \cdots\right), \\
x_3 &= -\frac{1-\mu}{2}(1+\mu)^{-2/3} - \frac{9+14\mu+13\mu^2}{12\mu}(1+\mu)^{-5/3}\sigma_1 + \cdots \\
&\quad + \left\{\left(\frac{3^{1/2}}{27} + \frac{5}{12}\right)\frac{1}{\mu} - \left(\frac{2(3)^{1/2}}{81} - \frac{29}{36}\right) - \left(\frac{3^{1/2}}{81} + \frac{7}{9}\right)\mu\right. \\
&\quad \left. + \cdots + \left[\left(\frac{5(3)^{1/2}}{2\cdot 3^5} - \frac{1687}{2\cdot 3^6}\right)\frac{1}{\mu^2} + \cdots\right]\sigma_1 + \cdots\right\}\sigma_2 + \cdots, \\
y_3 &= \frac{3^{1/2}}{2}\left((1+\mu)^{1/3} - \frac{3+\mu}{6\mu}(1+\mu)^{-2/3}\sigma_1 + \cdots\right. \\
&\quad + \left\{\left(\frac{2(3)^{1/2}}{81} + \frac{5}{18}\right)\frac{1}{\mu} + \left(\frac{8(3)^{1/2}}{243} + \frac{11}{54}\right)\right. \\
&\quad \left. + \left(\frac{2(3)^{1/2}}{81} - \frac{8}{27}\right)\mu + \cdots\right. \\
&\quad \left.\left. + \left[\left(\frac{41(3)^{1/2}}{3^6} + \frac{3173}{2^23^7}\right)\frac{1}{\mu^2} + \cdots\right]\sigma_1 + \cdots\right\}\sigma_2 + \cdots\right),
\end{aligned}
\tag{14.5}
$$

$$
\begin{aligned}
x_4 = &-\frac{1-\mu}{2}(1+\mu)^{-2/3}+\left[\left(\frac{3^{1/2}}{27}+\frac{5}{12}\right)\frac{1}{\mu}-\left(\frac{2(3)^{1/2}}{81}-\frac{29}{36}\right)\right.\\
&\left.-\left(\frac{3^{1/2}}{81}+\frac{7}{9}\right)\mu+\cdots\right]\sigma_1+\cdots\\
&+\left\{-\frac{9+14\mu+13\mu^2}{12\mu}(1+\mu)^{-5/3}\right.\\
&\left.+\left[\left(\frac{5(3)^{1/2}}{2\cdot 3^5}-\frac{1687}{2\cdot 3^6}\right)\frac{1}{\mu^2}+\cdots\right]\sigma_1+\cdots\right\}\sigma_2+\cdots,\qquad (14.5)\\
y_4 = &-\frac{3^{1/2}}{2}\left((1+\mu)^{1/3}+\left[\left(\frac{2(3)^{1/2}}{81}+\frac{5}{18}\right)\frac{1}{\mu}+\left(\frac{8(3)^{1/2}}{243}+\frac{11}{54}\right)\right.\right.\\
&\left.+\left(\frac{2(3)^{1/2}}{81}-\frac{8}{27}\right)\mu+\cdots\right]\sigma_1+\cdots+\left\{-\frac{3+\mu}{6\mu}(1+\mu)^{-2/3}\right.\\
&\left.\left.+\left[\left(\frac{41(3)^{1/2}}{3^6}+\frac{3173}{2^2\cdot 3^7}\right)\frac{1}{\mu^2}+\cdots\right]\sigma_1+\cdots\right\}\sigma_2+\cdots\right).
\end{aligned}
$$

15. **Points of libration for $n=4, 5$.** In order to find a point where $m_5=0$ may be introduced, we must solve the pair of equations given by the last of equations (13.2) for the case $n=5$, namely,

$$
\begin{aligned}
&\frac{A_{51}}{(r_{51})^3}+\frac{A_{52}}{(r_{52})^3}\mu+\frac{A_{53}}{(r_{53})^3}\sigma_1+\frac{A_{54}}{(r_{54})^3}\sigma_2=0,\\
&-y_5+\frac{y_5}{(r_{51})^3}+\frac{y_5-y_2}{(r_{52})^3}\mu+\frac{y_5-y_3}{(r_{53})^3}\sigma_1+\frac{y_5-y_4}{(r_{54})^3}\sigma_2=0,
\end{aligned}\qquad (15.1)
$$

where $A_{ij}=x_iy_j-x_jy_i$.

If $\sigma_1=0$ these equations are precisely those which determine P_3, hence P_5 coincides with P_3 for $\sigma_1=0$. Furthermore, $P_5\equiv P_3$ is a solution of equations (15.1) for $\sigma_1\neq 0$, the identity being in μ, σ_1 and σ_2. In order to obtain a solution of (15.1) having the property (in §5) that $P_5\neq P_3$, we let $\sigma_1=3\nu^3$. Upon substituting

$$
\begin{aligned}
x_5 = &-\frac{1-\mu}{2}(1+\mu)^{-2/3}+p_1(\mu)\nu+p_2(\mu)\nu^2\\
&+p_3(\mu)\nu^3+\cdots+c_1(\mu)\sigma_2+\cdots,\qquad (15.2)\\
y_5 = &\frac{3^{1/2}}{2}\left[(1+\mu)^{1/3}+q_1(\mu)\nu+q_2(\mu)\nu^2\right.\\
&\left.+q_3(\mu)\nu^3+\cdots+c_2(\mu)\sigma_2+\cdots\right]
\end{aligned}
$$

in (15.1) the $p_i(\mu)$ and $q_i(\mu)$ may be determined by equating coefficients of ν.

The functions $c_1(\mu)$ and $c_2(\mu)$ are the same as in (14.4). The result is

$$p_1(\mu) = \pm\left(\frac{1}{2} - \frac{1}{4}\mu + \cdots\right), \quad p_2(\mu) = \pm\left(\frac{1}{6} + \frac{125}{144}\mu + \cdots\right),$$
$$p_3(\mu) = -\frac{9 + 14\mu + 13\mu^2}{4\mu}(1+\mu)^{-5/3},$$
$$q_1(\mu) = \pm\left(-1 - \frac{1}{2}\mu + \cdots\right), \quad q_2(\mu) = \pm\left(\frac{1}{3} + \frac{25}{72}\mu + \cdots\right), \tag{15.3}$$
$$q_3(\mu) = -\frac{3+\mu}{2\mu}(1+\mu)^{-2/3}.$$

Thus, as σ_1 increases from zero to a positive value, a point of libration, denoted by P_5^*, branches off from P_3 and moves to a finite distance from P_3. We shall use only the upper sign in (15.3).

The solution of

$$\frac{A_{61}}{(r_{61})^3} + \frac{A_{62}}{(r_{62})^3}\mu + \frac{A_{63}}{(r_{63})^3}\sigma_1 + \frac{A_{64}}{(r_{64})^3}\sigma_2 = 0,$$
$$-y_6 + \frac{y_6}{(r_{61})^3} + \frac{y_6 - y_2}{(r_{62})^3}\mu + \frac{y_6 - y_2}{(r_{63})^3}\sigma_1 + \frac{y_6 - y_4}{(r_{64})^3}\sigma_2 = 0 \tag{15.4}$$

for the libration point P_6^* may now be obtained from (15.2) by the transformation

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & y_2 & y_3 & y_4 & y_5 & y_6 & \mu & \sigma_1 & \sigma_2 \\ x_1 & x_2 & x_4 & x_3 & x_6 & x_5 & -y_2 & -y_4 & -y_3 & -y_6 & -y_5 & \mu & \sigma_2 & \sigma_1 \end{pmatrix},$$

which reduces (15.4) to (15.1). Therefore, as σ_2 increases from zero to a positive value, a point of libration P_6^*, whose coordinates are

$$x_6 = -\frac{1-\mu}{2}(1+\mu)^{-2/3} + c_1(\mu)\sigma_1 + \cdots + p_1(\mu)\left(\frac{\sigma_2}{3}\right)^{1/3}$$
$$+ p_2(\mu)\left(\frac{\sigma_2}{3}\right)^{2/3} + p_3(\mu)\left(\frac{\sigma_2}{3}\right) + \cdots,$$
$$y_6 = -\frac{3^{1/2}}{2}\left[(1+\mu)^{1/3} + c_2(\mu)\sigma_1 + \cdots + q_1(\mu)\left(\frac{\sigma_2}{3}\right)^{1/3}\right. \tag{15.5}$$
$$\left. + q_2(\mu)\left(\frac{\sigma_2}{3}\right)^{2/3} + q_3(\mu)\left(\frac{\sigma_2}{3}\right) + \cdots\right],$$

branches off from P_4 and moves to a finite distance from P_4.

As the masses m_5 and m_6 increase from zero to positive values, the solution

functions (14.5), (15.2) and (15.5) must be extended to include powers of $(\sigma_3/3)^{1/3}$ and $(\sigma_4/3)^{1/3}$. This can readily be accomplished by the substitution

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & y_2 & y_3 & y_4 & y_5 & y_6 & \mu & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ x_1 & x_2 & x_5 & x_6 & x_3 & x_4 & y_2 & y_5 & y_6 & y_3 & y_4 & \mu & \sigma_3 & \sigma_4 & \sigma_1 & \sigma_2 \end{pmatrix},$$

which leaves the system (13.1), for $n=6$, unchanged.

16. **The solution functions for any n.** It is evident that the process carried out in the last section may be repeated until the system contains any finite number of masses. Concerning the points of libration at any stage, we have the following theorem.

THEOREM 16.1. *As masses m_{2k-1} and m_{2k} are added to the system, by allowing them to increase from zero to some sufficiently small positive values at points P_{2k-1} and P_{2k} respectively, two points of libration P^*_{2k+1} and P^*_{2k+2} branch off from P_{2k-1} and P_{2k} respectively.*

In regard to the solution functions at any stage the following theorem is evident.

THEOREM 16.2. *As masses m_{2k+1} and m_{2k+2} increase from zero to some sufficiently small positive values, the solution functions x_i and y_i $(i=1, 2, \cdots, 2k+2)$ vary continuously and may be expressed as power series in $(\sigma_{2k-1}/3)^{1/3}$ and $(\sigma_{2k}/3)^{1/3}$, where σ_{2k-1} and σ_{2k} are the mass ratios m_{2k+1}/m_1 and m_{2k+2}/m_1 respectively.*

As pointed out in §14, all the solution functions for $n=4$ are power series in σ_1 and σ_2. This is due to the fact that equations (13.1) are invariant under the transformations

$$\begin{pmatrix} x_i & y_i & \sigma_{i-2} \\ x_j & \pm y_j & \sigma_{j-2} \end{pmatrix}, \qquad i, j = 3, 4, \cdots, n,$$

where the sign is to be taken "+" if i, j are both even or both odd, otherwise "−," and also to the fact that the solution functions if $n=4$ are power series in σ_1. For $n>4$ we note that only x_1, x_2 and y_2 are power series in σ_1, and hence only these three functions are power series in the mass ratios $\sigma_1, \sigma_2, \cdots, \sigma_{n-2}$. Since x_5 and y_5 are power series in $(\sigma_1/3)^{1/3}$, it follows that x_i and y_i $(i=3, 4, \cdots, n)$ are power series in $(\sigma_1/3)^{1/3}$, $(\sigma_2/3)^{1/3}$, $\cdots$, $(\sigma_{n-2}/3)^{1/3}$. It should be noted, however, that x_i and y_i $(i=3, 4, \cdots, n)$ may be power series in $(\sigma_1/3)^{1/3}$, $(\sigma_2/3)^{1/5}$, $\cdots$, $(\sigma_{n-3}/3)^{1/3}$, $(\sigma_{n-2}/3)$, but this solution does not have the property stated in Theorem 5.1, for as m_{n-2} vanishes P_{n-1} and P_n coincide.

Let $(\sigma_i/3)^{1/3}=\tau_i^{1/3}$. In case the solution functions are power series in $\tau_i^{1/3}$, $i=1, 2, \cdots, n-2$, the following functions have all the properties stated in theorems of §§5, 6, 7, 8 and constitute a solution of (13.1).

$$
\begin{aligned}
x_1 &= H_{1,0} + H_{1,1}\sum \tau_{2j-1}^{1/3} + \overline{H}_{1,1}\sum \tau_{2j}^{1/3} + H_{1,2}\sum \tau_{2j-1}^{2/3} + \overline{H}_{1,2}\sum \tau_{2j}^{2/3} \\
&\quad + H_{1,3}\sum \tau_{2j-1} + \overline{H}_{1,3}\sum \tau_{2j} + \cdots, \\
y_1 &= K_{1,0} + K_{1,1}\sum \tau_{2j-1}^{1/3} + \overline{K}_{1,1}\sum \tau_{2j}^{1/3} + K_{1,2}\sum \tau_{2j-1}^{2/3} + \overline{K}_{1,2}\sum \tau_{2j}^{2/3} \\
&\quad + K_{1,3}\sum \tau_{2j-1} + \overline{K}_{1,3}\sum \tau_{2j} + \cdots, \\
x_2 &= H_{2,0} + H_{2,1}\sum \tau_{2j-1}^{1/3} + \overline{H}_{2,1}\sum \tau_{2j}^{1/3} + H_{2,2}\sum \tau_{2j-1}^{2/3} + \overline{H}_{2,2}\sum \tau_{2j}^{2/3} \\
&\quad + H_{2,3}\sum \tau_{2j-1} + \overline{H}_{2,3}\sum \tau_{2j} + \cdots, \\
y_2 &= K_{2,0} + K_{2,1}\sum \tau_{2j-1}^{1/3} + \overline{K}_{2,1}\sum \tau_{2j}^{1/3} + K_{2,2}\sum \tau_{2j-1}^{2/3} + \overline{K}_{2,2}\sum \tau_{2j}^{2/3} \\
&\quad + K_{2,3}\sum \tau_{2j-1} + \overline{K}_{2,3}\sum \tau_{2j} + \cdots, \qquad (16.1) \\
x_{2i+1} &= H_{2i+1,0} + H_{2i+1,1}\sum{}'\tau_{2j-1}^{1/3} + \overline{H}_{2i+1,1}\sum{}'\tau_{2j}^{1/3} + H_{2i+1,2}\sum{}'\tau_{2j-1}^{2/3} \\
&\quad + \overline{H}_{2i+1,2}\sum{}'\tau_{2j}^{2/3} + H_{2i+1,3}\sum \tau_{2j-1} + \overline{H}_{2i+1,3}\sum \tau_{2j} + \cdots, \\
y_{2i+1} &= K_{2i+1,0} + K_{2i+1,1}\sum{}'\tau_{2j-1}^{1/3} + \overline{K}_{2i+1,1}\sum{}'\tau_{2j}^{1/3} + K_{2i+1,2}\sum{}'\tau_{2j-1}^{2/3} \\
&\quad + \overline{K}_{2i+1,2}\sum{}'\tau_{2j}^{2/3} + K_{2i+1,3}\sum \tau_{2j-1} + \overline{K}_{2i+1,3}\sum \tau_{2j} + \cdots, \\
x_{2i+2} &= H_{2i+2,0} + H_{2i+2,1}\sum{}'\tau_{2j-1}^{1/3} + \overline{H}_{2i+2,1}\sum{}'\tau_{2j}^{1/3} + H_{2i+2,2}\sum{}'\tau_{2j-1}^{2/3} \\
&\quad + \overline{H}_{2i+2,2}\sum{}'\tau_{2j}^{2/3} + H_{2i+2,3}\sum \tau_{2j-1} + \overline{H}_{2i+2,3}\sum \tau_{2j} + \cdots, \\
y_{2i+2} &= K_{2i+2,0} + K_{2i+2,1}\sum{}'\tau_{2j-1}^{1/3} + \overline{K}_{2i+2,1}\sum{}'\tau_{2j}^{1/3} + K_{2i+2,2}\sum{}'\tau_{2j-1}^{2/3} \\
&\quad + \overline{K}_{2i+2,2}\sum{}'\tau_{2j}^{2/3} + K_{2i+2,3}\sum \tau_{2j-1} + \overline{K}_{2i+2,3}\sum \tau_{2j} + \cdots,
\end{aligned}
$$

where $\sum$ denotes the sum over j from 1 to the largest integer in $(n-2)/2$, and $\sum'$ the same sum over j except for $j=i$. The H's and K's are functions of μ and the nonzero ones are

$$
\begin{aligned}
&H_{1,0} = \mu(1+\mu)^{-2/3}, && H_{1,2} = \overline{H}_{1,2} = \frac{3-\mu}{2}(1+\mu)^{-5/3}, \\
&H_{2,0} = -(1+\mu)^{-2/3}, \\
&H_{2,2} = \overline{H}_{2,2} = \frac{1-3\mu}{2}(1+\mu)^{-5/3}, && K_{2,2} = -\overline{K}_{2,2} = -\frac{3}{\mu}(1+\mu)^{1/3}, \\
&H_{2i+1,0} = H_{2i+2,0} = -\frac{1-\mu}{2}(1+\mu)^{-2/3}, \\
&K_{2i+1,0} = -K_{2i+2,0} = \frac{3^{1/2}}{2}(1+\mu)^{1/3},
\end{aligned}
$$

$$H_{2i+1,1} = \overline{H}_{2i+2,1} = \frac{1}{2} - \frac{1}{4}\mu + \cdots,$$

$$K_{2i+1,1} = -\overline{K}_{2i+2,1} = \frac{3^{1/2}}{2}\left(-1 - \frac{1}{2}\mu + \cdots\right),$$

$$H_{2i+1,2} = \overline{H}_{2i+2,2} = -\frac{1}{6} + \frac{125}{144}\mu + \cdots,$$

$$K_{2i+1,2} = -\overline{K}_{2i+2,2} = \frac{3^{1/2}}{2}\left(\frac{1}{3} + \frac{25}{72}\mu + \cdots\right),$$

$$H_{2i+1,3} = \overline{H}_{2i+2,3} = -\frac{9 + 14\mu + 13\mu^2}{4\mu}(1+\mu)^{-5/3},$$

$$K_{2i+2,3} = -\overline{K}_{2i+2,3} = -\frac{3^{1/2}\ (3+\mu)}{4\mu}(1+\mu)^{-2/3},$$

$$H_{2i+2,3} = \overline{H}_{2i+1,3} = \frac{3^{1/2}}{9\mu} + \frac{5}{4\mu} - \frac{2(3)^{1/2}}{27} + \frac{29}{12} - \frac{3^{1/2}}{27}\mu - \frac{7}{3}\mu + \cdots,$$

$$K_{2i+2,3} = -\overline{K}_{2i+1,3}$$

$$= -\frac{3^{1/2}}{2}\left(\frac{2(3)^{1/2}}{27\mu} + \frac{5}{6\mu} + \frac{8(3)^{1/2}}{81} + \frac{11}{18} + \frac{2(3)^{1/2}}{27}\mu - \frac{8}{9}\mu + \cdots\right).$$

Since (16.1) is a solution of (13.1), and hence of (2.1), for values of $\sigma_1, \sigma_2, \cdots, \sigma_{n-2}$ for which the series converge, and these values may be attained by choosing m_1 sufficiently large, we may state Theorem 12.2 as follows:

THEOREM 16.3. *Given $n-1$ arbitrary finite masses it is possible to choose an additional mass sufficiently large and an angular velocity sufficiently great, so that the n masses form a noncollinear plane permanent configuration.*

17. **The polygon formed by the P_i.** A diagram illustrates clearly the properties of a solution. In case $n=1$ the polygon, Fig. 3a, consists of one point P_1 at the origin. There is one point of libration P_2^* at a distance $(\lambda/m_1)^{1/3}=1$ from P_1. As a zero mass m_2 is placed at P_2^* and allowed to increase to a positive value, so that $(m_2/m_1)=\mu>0$, the configuration for $n=2$ is obtained. The polygon now consists of the line segment P_1P_2, Fig. 3b, with two points of libration P_3^* and P_4^* not collinear(7) with P_1 and P_2. These points are the vertices of equilateral triangles with side $P_1P_2=(1+\mu)^{1/3}$.

As a zero mass m_3 is placed at P_3^* and allowed to increase to a posi-

(7) The points of libration collinear with P_1P_2 are obtained by solving the first three equations (2.1) for x_3, where $n=3$ and $y_1=y_2=y_3=m_3=0$. For fixed order x_1, x_2, x_3 this is equivalent to Lagrange's quintic, cf. Lagrange [8, p. 277], or Tisserand [15, p. 155].

tive value, so that $(m_3/m_1)=\sigma_1>0$, the configuration for $n=3$ is obtained. The polygon in this case is the triangle $P_1P_2P_3$, Fig. 3c, which by the Theorem of Lagrange is an equilateral triangle of side $(1+\mu+\sigma_1)^{1/3}$. There are at least two(8) points of libration P_4^* and P_5^*. In general there is an $(n+2)$-sided polygon consisting of n points P_i and two libration points P_{n+1}^* and P_{n+2}^*.

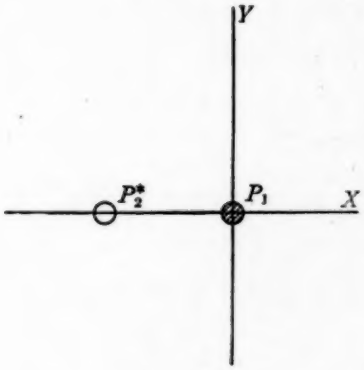


FIG. 3a

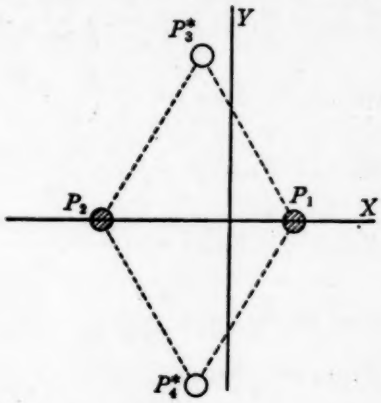


FIG. 3b

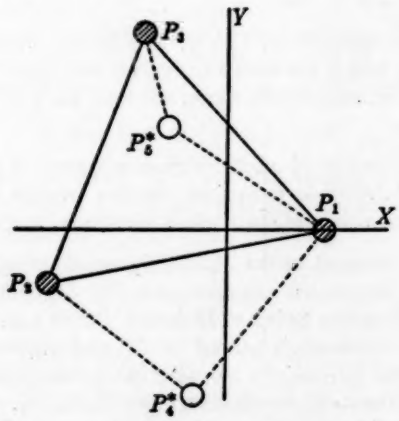


FIG. 3c

From the solution functions (16.1) we observe an interesting property which we state in the following theorem.

(8) According to Henrichsen [6], if the three masses are equal there are ten points of libration.

THEOREM 17.1. *As the masses m_i $(i=3, 4, \cdots, n)$, located respectively at points P_i, approach zero the points P_i cluster about the two libration points for the two body problem.*

The geometric properties of these n-sided polygons, whether they are convex or concave, have not been determined. Furthermore, the possibility of removing the restrictions on the mass ratios seems to depend upon whether the determinant D_n in §12 is different from zero for all values of the masses.

BIBLIOGRAPHY

1. H. Andoyer, *Sur l'équilibre relatif de n-corps*, Bulletin Astronomique vol. 23 (1906) pp. 50–59.

2. Emilia Breglia, *Su alcuni casi particolari del problema dei tre corpi*, Giornale de Mathematiche (3) vol. 7 (1916) pp. 151–173.

3. O. Dziobek, *Über einen merkwürdigen Fall des Vielkörperproblems*, Astronomische Nachrichten vol. 152 (1900) pp. 33–46.

4. L. Euler, *Theoria motuum planetarum et cometarum*, 1744.

5. L. Euler, *De motu rectilineo trium corporum se mutuo attrahentium*, Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae vol. 11 (1776); *Considérations sur le problème des trois corps*, Mémoirs de l'Académie des Sciences de Berlin 1763 (1770) pp. 194–220.

6. J. J. L. Hinrichsen, *Bounds for the libration points in a restricted problem of n-bodies*, Bull. Amer. Math. Soc. vol. 48 (1942) p. 45.

7. R. Hoppe, *Erweiterung der bekannten Speziallösung des dreikörper Problems*, Archiv der Mathematik und Physik vol. 64 (1879) pp. 218–233.

8. J. L. Lagrange, *Essai sur le problème des trois corps*, Œuvres vol. 6 (1772) pp. 229–324.

9. R. Lehmann-Filhès, *Über zwei Fälle des vielkörper Problems*, Astronomische Nachrichten vol. 127 (1891) pp. 137–144.

10. W. R. Longley, *Some particular solutions in the problem of n-bodies*, Bull. Amer. Math. Soc. vol. 13 (1906–1907) pp. 324–335.

11. W. D. MacMillan, *Dynamics of rigid bodies*, 1936.

12. W. D. MacMillan and W. Bartky, *Permanent configurations in the problem of four bodies*, Trans. Amer. Math. Soc. vol. 34 (1932) pp. 838–875.

13. F. R. Moulton, *Straight line solutions in the problem of n-bodies*, Ann. of Math. (2) vol. 12 (1910) pp. 1–17.

14. Newton, *Principia*, book one, §2, 3d ed., 1725.

15. F. Tisserand, *Mécanique céleste*, vol. 1, 1889, chap. 8.

16. W. L. Williams, *Permanent configurations in the problem of five bodies*, Trans. Amer. Math. Soc. vol. 44 (1938) pp. 562–579.

17. A. Wintner, *Analytical foundation of celestial mechanics*, 1941.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.